
بسم الله الرحمن الرحيم

دانشکده مهندسی برق

پایان‌نامه کارشناسی ارشد مهندسی مخابرات سیستم

فراتفکیک پذیری در ویدیوی رنگی *HD* با استفاده از شبکه های عمیق

نگارنده: علی مظفری

استاد راهنما

دکتر حسین خسروی

دی ۱۳۹۹

شماره: ۱۷۸۰/آ.ت.ب
تاریخ: ۹۹/۱/۲۴

باسمه‌تعالی

دانشگاه صنعتی شاهرود
مدیریت تحصیلات تکمیلی

فرم شماره (۳) صورتجلسه نهایی دفاع از پایان نامه دوره کارشناسی ارشد

با نام و یاد خداوند متعال، ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد آقای **علی مظفری** با شماره دانشجویی **۹۵۱۴۰۱۴** رشته **مهندسی برق** گرایش **(مخابرات سیستم)** تحت عنوان : **فراتفکیک پذیری ویدیوی رنگی HD با استفاده از شبکه های عمیق** که درتاریخ **۱۳۹۹/۱۰/۲۴** با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می گردد:

الف) درجه عالی: نمره ۲۰-۱۹ ☐ ب) درجه خیلی خوب: نمره ۱۸/۹۹ – ۱۸ ☑
ج) درجه خوب: نمره۱۷/۹۹– ۱۶ ☐ د) درجه متوسط: نمره ۱۵/۹۹-۱۴ ☐
ه) کمتر از ۱۴ غیر قابل قبول و نیاز به دفاع مجدد دارد ☐
نوع تحقیق: نظری ☑ **عملی** ☐

عضو هیأت داوران	نام ونام خانوادگی	مرتبهٔ علمی	امضاء
۱ـ استادراهنمای اول	دکتر حسین خسروی	استادیار	
۲- استادراهنمای دوم	—	—	
۳- استاد مشاور	—	—	
۴- نماینده تحصیلات تکمیلی	دکتر امیدرضا معروضی	استادیار	
۵- استاد ممتحن اول	دکتر علیرضا احمدی فرد	دانشیار	
۶- استاد ممتحن دوم	دکتر هادی گرایلو	استادیار	

نام و نام خانوادگی رئیس دانشکده: دکتر علیرضا احمدی فرد

تاریخ و امضاء و مهر دانشکده:

تبصره: در صورتی که کسی مردود شود حداکثر یکبار دیگر (در مدت زمان تحصیل) می تواند از پایان نامه خود دفاع نماید (دفاع مجدد نباید زودتر از ۴ ماه برگزار شود).

این پایان نامه را در کمال افتخار و امتنان تقدیم می نمایم به

محضر ارزشمند پدر و مادر دلسوز و مهربانم

به پاس تعبیر عظیم و انسانی شان از کلمه ایثار و از خودگذشتگی

به پاس گرمای امیدبخش وجودشان که همواره موجب آرامش روحی و آسایش فکری من بوده

و به پاس حمایت های همه جانبه خود که همواره کننده مسیر زندگی من شدند.

تشکر و قدردانی

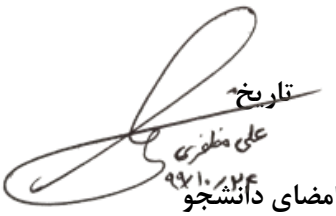
شکر شایان نثار ایزد منان که آثار قدرت او بر چهره روز روشن، تابان است و انوار حکمت او در دل شب تار، درفشان. آفریدگاری که خویشتن را به ما شناساند و درهای علم را بر ما گشود و عمری و فرصتی عطا فرمود تا بدان، بنده ضعیف خویش را در طریق علم و معرفت بیازماید.

همچنین از استاد گرامی، جناب آقای دکتر حسین خسروی که در کمال سعه صدر با حسن خلق و فروتنی از هیچ کمکی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این پایان نامه را در حالی متقبل شدند که بدون مساعدت ایشان، این پژوهش هرگز به نتیجه نمی رسید.

تعهد نامه

اینجانب **علی مظفری** دانشجوی دوره کارشناسی ارشد رشته **مهندسی برق** دانشکده مهندسی برق دانشگاه صنعتی شاهرود نویسنده پایان‌نامه **فراتفکیک پذیری در ویدیوی *HD* با استفاده از شبکه های عمیق** تحت راهنمائی **دکتر حسین خسروی** متعهد می‌شوم.

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است .
- در استفاده از نتایج پژوهشهای محققان دیگر به مرجع مورد استفاده استناد شده است .
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « ***Shahrood University of Technology*** » به چاپ خواهد رسید .
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده اند در مقالات مستخرج از پایان‌نامه رعایت می گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است .
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است .



امضای دانشجو

مالکیت نتایج و حق نشر

کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج ، کتاب ، برنامه های رایانه ای ، نرم افزار ها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می‌باشد . این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود .

استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

چکیده

امروزه کیفیت تصویر و ویدیو از اهمیت ویژه ای برخوردار است. تکنیک فراتفکیک پذیری از روش های نرم افزاری در حوزه پردازش سیگنال است که به منظور افزایش کیفیت تصویر از لحاظ افزایش تعداد پیکسل ها و کاهش نویز، استفاده می شود. در این تکنیک، با استفاده از یک و یا چند تصویر کم وضوح، تصویری با وضوح بالاتر ایجاد می شود.

در این پایان نامه یک روش فراتفکیک پذیری ویدیو مبتنی بر یادگیری و بر اساس شبکه های همگشتی ارائه می کنیم که با استفاده از یک شبکه آموزش پذیر، شار نوری (*Optical Flow*) موجود بین فریم کنونی و قبلی را محاسبه نموده و به عنوان اطلاعات اضافی به فریم کنونی وضوح پایین اضافه می کنیم. این اطلاعات به سامانه‌ی فراتفکیک پذیری داده شده و سامانه‌ی مذکور، با استفاده از شبکه های همگشتی عمیق به بازسازی تصویر می پردازد. در نهایت تصویر بازسازی شده به عنوان بلوک مولد یک شبکه *GAN* به بلوک متمایزگر داده شده که تصویری با وضوح بالا و واقعی را نتیجه می دهد.

روش پیشنهادی با استفاده از سه بلوک آموزش پذیر و با به کارگیری شبکه های همگشتی عمیق، دریافت اطلاعات حاصل از شار نوری و همچنین بهره گیری از شبکه های *GAN*، توانسته به بزرگنمایی ویدیو تا ۴ برابر و بدون افت کیفیت محسوس دست پیدا کند.

در انتها مدل پیشنهادی (*FRGAN-VSR*) جهت ارزیابی کیفیت با دو مدل *SRGAN* (حذف شبکه شار نوری از مدل پیشنهادی) و *FRVSR* (حذف شبکه *GAN* از مدل پیشنهادی) در سه فاز یادگیری، ارزیابی و آزمایش مقایسه شده است. جهت بررسی کیفیت بصری، مدل پیشنهادی برای مجموعه داده *Vid4* توسط دو معیار *PSNR* و *SSIM* مورد بررسی قرار گرفتند. همچنین کیفیت بصری مدل در چندین فریم نمونه از یک ویدیو و نیز چند ویدیوی چند فریمی مورد بررسی قرار گرفت. علاوه بر این در آزمایش پیوستگی زمانی مدل ها را

جهت بررسی حفظ پیوستگی فریم ها در فراتفکیک پذیری ویدیو مورد مقایسه قرار دادیم، نتایج به دست آمده نشان می دهد که روش پیشنهادی از تمامی روش های مشابه حال حاضر چه از لحاظ کیفی، چه کمی و چه پیوستگی زمانی نتایج بهتری را از خود بر جای می گذارد.

کلمات کلیدی: فراتفکیک پذیری، ویدیو، وضوح، شبکه های عصبی همگشتی، شار نوری، *GAN*

فهرست مطالب

فهرست جداول

فهرست اشکال

فصل ۱
مقدمه

۱ – ۱ – مقدمه

در سال های اخیر، پیشرفت های گسترده ای در زمینه ی حسگر[1]های تصویر و سیستم های تصویربرداری دیجیتال صورت گرفته است. اما همچنان محدودیت های تئوری و عملی جهت داشتن تصویر و ویدیو با تفکیک پذیری بالا[2] وجود دارد. تمایل برای استفاده از تصاویری با وضوح بالا از دو زمینه اصلی نشات می گیرد: بهبود اطلاعات تصویری برای تفسیر انسان و کمک به درک دستگاه های خودکار. لذا تلاش های بسیاری جهت افزایش وضوح تصاویر و ویدیوهای دیجیتالی صورت گرفته است که به دو دسته کلی روش های سخت افزاری و نرم افزاری تقسیم بندی می شود.

در روش های سخت افزاری، علی رغم غیرقابل انکار بودن پیشرفت های چشم گیر در تولید دوربین های دیجیتال با وضوح بالا، به دلیل محدودیت های تکنیکی موجود در تکنولوژی ساخت مدارات مجتمع، رسیدن به تصاویری با کیفیت و وضوح بالاتر، به حسگرهای تصویر محدود می شود. ضمن آنکه کاربردهای خاص نظیر استفاده از دوربین در ماهواره های فضایی و یا نیاز به چیپ های تا حد امکان کوچک در تلفن های همراه عاملی محدود کننده و بسیار تاثیرگذار در بهبود کیفیت تصویر در روش های سخت افزاری می باشد. مهم تر از همه آنکه ارتقا کیفیت تصویر در این حوزه بسیار پرهزینه است.

در مقابل، پیشرفت قابل توجه پردازشگرهای رایانه ای موجب کاهش هزینه‌های محاسباتی شده است. سهولت استفاده از روش های نرم افزاری، عدم محدودیت های فیزیکی و هزینه‌ی بسیار پایین، روش های نرم افزاری بهبود کیفیت تصویر را به گزینه ای بسیار رقابتی و جذاب در برابر روش های سخت افزاری تبدیل کرده است.

تکنیک های فراتفکیک پذیری[3] به عنوان گونه ای از روش های نرم افزاری در حوزه پردازش سیگنال و به منظور غلبه بر محدودیت های مذکور معرفی می شوند. عنوان فراتفکیک پذیری به دلیل اینکه قادر خواهیم

---

[1] Sensor
[2] High Resolution
3 Super Resolution

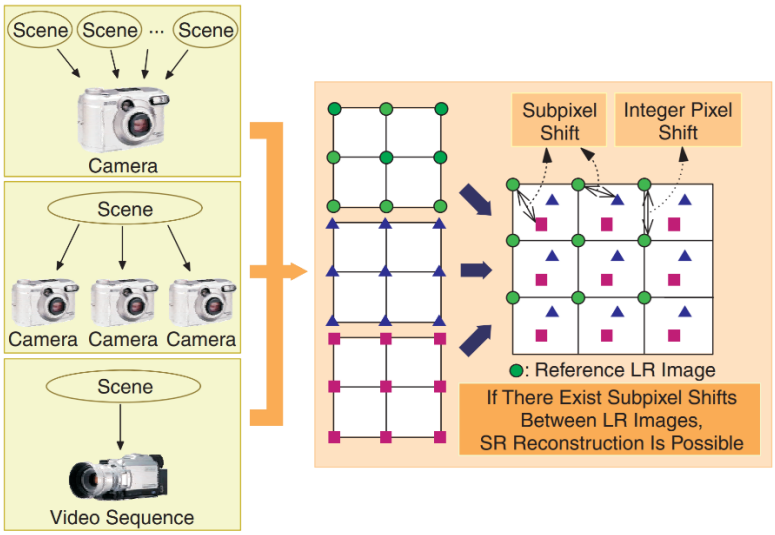
بود که از محدوده توانایی سیستم تصویربرداری فراتر رویم بدین نام، نامگذاری شده است. مطالعات و پیشرفت های فراوانی به خصوص در سال های اخیر در این حوزه انجام شده است. این تکنیک ها با استفاده از یک و یا چند تصویر با وضوح پایین[1] تصویری با وضوح بالاتر[2] ایجاد می کنند. پژوهش های اخیر در زمینه ی فراتفکیک پذیری، به منظور کاهش پیچیدگی محاسباتی و افزایش مقاومت در برابر خطاهای مدل سازی و نویز ایجاد شده است.

۱ – ۲ – شرح مسئله

فراتفکیک پذیری به دسته ای از روش های پردازش سیگنال اطلاق می شود که با دریافت یک یا چند تصویر با وضوح پایین قادر به بازیابی یک تصویر با وضوح بالا باشد. بسیاری از روشهای فراتفکیک‌پذیری بر مبنای یک ایده می باشند. استفاده از اطلاعات چندین فریم از یک تصویر که حاوی اطلاعات منحصر به خود بوده و ترکیب این اطلاعات برای بازیابی یک تصویر با وضوح و جزئیات بیشتر [۱] .

فریم های با وضوح پایین می توانند با استفاده از حرکت یک دوربین و یا با جایگذاری چندین دوربین در مکان های مناسب فراهم گردند. به دلیل آنکه نیازمند استخراج اطلاعات منحصر به فرد از هر فریم هستیم، ضروریست که فریم های تهیه شده دارای جابه جایی *sub-pixel* نسبت به یکدیگر باشند. در غیراینصورت پیکسل های فریم های گوناگون بر هم منطبق شده و اطلاعات و جزئیات بیشتری را نسبت به تصویر با وضوح پایین نخواهد داد. لذا بازیابی تصویر با وضوح بالا امکان پذیر نخواهد بود. شکل ۱-۱ این الگوریتم را به صورت کامل نشان می دهد.

1 Low Resolution (LR)
2 High Resolution (HR)



شکل ۱-۱- نمونه ای از فریم های تهیه شده از یک صحنه که نسبت به همدیگر شیفت صحیح ندارند [۱]

با ترکیب فریم های تهیه شده با شرایط مذکور و نگاشت آن به چهارچوب شبکه[1] تصویر با وضوح بالا می توان به یک تصویر بزرگتر و با کیفیت بالاتر دست یافت.

لازم به ذکر است که تعریف بیان شده از فراتفکیک پذیری را می توان به نوعی یک تعریف کلاسیک از این مبحث به حساب آورد که چندان دقیق نبوده و یا به عبارت دیگر نمی توان آن را مبنای کلیه روش های موجود در پژوهش های فراتفکیک پذیری به حساب آورد. توضیح آنکه روش بیان شده اگرچه اطلاعات بسیار زیادی در تولید یک تصویر با وضوح بالا به ما خواهد داد اما تهیه چندین فریم از یک صحنه با شرایط گفته شده عمدتاً امکان پذیر نبوده و یا بسیار دشوار است. لذا بسیاری از روش های مدرن فراتفکیک پذیری تصویر به دنبال راه حل هایی هستند که به چندین فریم ورودی نیاز نداشته و بتوان از یک فریم ورودی با وضوح پایین، یک خروجی با وضوح بالا را به دست آورد. الگوریتم های فراتفکیک پذیری تک فریمی می توانند به چهاردسته

[1] Grid

مدل های پیش‌بین[1] ، روش های مبتنی بر لبه[2]، روش های آماری تصویر[3] و مبتنی بر نمونه[4] دسته بندی نمود که از این بین، روش های مبتنی بر نمونه از عملکرد بالاتری برخوردار هستند [۲].

از سوی دیگر در فراتفکیک پذیری ویدیو به دلیل آنکه تمام فریم های ویدیو را به ترتیب در اختیار داریم، اگر بتوان راه حلی پیدا نمود که اطلاعات حاصل از فریم های مجاور را جهت بازسازی هر فریم به کمک فراتفکیک پذیری تک فریمی ادغام نمود می توان به دقت بالاتری در بازسازی تصویر دست پیدا کرد.

۱ – ۳ – اهمیت انجام پژوهش

روش های مبتنی بر فراتفکیک پذیری را می توان به گونه ای در تقابل با روش های کلاسیک بزرگنمایی تصویر نظیر روش های درون یابی در نظر گرفت. به گونه ای که جهت تولید تصویر بزرگنمایی شده به جای انجام محاسبات ریاضی همواره یکسان و بی ارتباط با ورودی، بتوان ارتباط معناداری بین ورودی ها پیدا کرده و از آن جهت تولید تصویر بزرگ نمایی شده استفاده نمود. نتیجه آنکه تصاویر فراتفکیک پذیر شده به طرز شگرفی از جزئیات بسیار بالاتری نسبت به روش های کلاسیک برخوردار خواهند بود. البته این سخن به معنای عدم کاربرد روش های کلاسیک نظیر درون یابی نیست. در تایید آن همین بس که در بخشی از فرآیند فراتفکیک پذیری از روش های درون یابی نظیر درون یابی دو خطی[5] استفاده می کنیم. در شکل ۱-۲ به مقایسه عملکرد بین روش های کلاسیک بزرگنمایی نظیر درون یابی و یکی از الگوریتم های فراتفکیک پذیری تک فریمی پرداخته ایم.

[1] Prediction Models
[2] Edge based methods
[3] Image statistical methods
[4] Example based methods
[5] Bilinear Interpolation

شکل ۲-۱۱- مقایسه عملکرد بزرگنمایی به روش درون یابی کلاسیک و بزرگنمایی به روش فراتفکیک پذیری

نتایج بسیار قابل توجه و عملکرد خیره کننده فراتفکیک پذیری نسبت به سایر روش های کلاسیک بزرگنمایی، ما را به کاربرد های متنوع فراتفکیک پذیری نظیر کاربردهای پزشکی [۳]، تصاویر ماهواره ای [۴] و کاربردهای نظارتی [۵] رهنمون می سازد. همچنین با ظهور نمایشگرهای $4K$ و $8K$ و عدم وجود محتوای سازگار با نمایشگرهای وضوح بالا، وجود راهکاری که بتواند ویدیو های وضوح پایین را با عملکردی مناسب و کیفیتی قابل قبول افزایش مقیاس دهد نیاز می شود.

۱ – ۴ – هدف پژوهش

هدف اصلی این پایان نامه، طراحی الگوریتمی است که بتواند وضوح یک ویدیو را افزایش دهد بدون آنکه تصویر دچار افت کیفیت محسوس گردد. روش پیشنهادی ما در زمره فراتفکیک پذیری های تک فریمی مبتنی بر یادگیری قرار می گیرد و نسبت به سایر روش های موجود از کیفیت و عملکرد بهتری برخوردار است. از آنجا که فراتفکیک پذیری تک فریمی و ترکیب تمام فریم های با وضوح بالا با یکدیگر، پیوستگی زمانی که در ویدیو با وضوح پایین داشت را تامین نمی کند؛ راه حلی ارائه خواهیم کرد که بتواند با استفاده از اطلاعات فریم های مجاور، پیوستگی فریم ها را در ویدیوی فراتفکیک پذیر شده نیز تامین نماید. استفاده از اطلاعات فریم های

مجاور موجب مشکل دیگری خواهد شد و آن افزایش میزان پیچیدگی محاسباتی الگوریتم می باشد. در این پژوهش سعی شده با در نظر گرفتن این مشکل به راه حلی جهت کاستن پیچیدگی محاسباتی بپردازیم.

۱ – ۵ – ساختار پایان نامه

این پایان نامه به ۵ فصل تقسیم شده است. پس از بیان مقدماتی که در این فصل بیان گردید در فصل ۲ به بیان کاربردهای فراتفکیک پذیری و پژوهش هایی که به طور تخصصی در کاربردی خاص از فراتفکیک پذیری استفاده نموده است خواهیم پرداخت. ضمن آنکه پژوهش های انجام گرفته در این حوزه را با در نظر گرفتن نوع الگوریتم تقسیم بندی خواهیم نمود. در نهایت با توجه به اینکه پایان نامه موجود یک روش فراتفکیک پذیری مبتنی بر یادگیری است و نمونه های پیشین را بسط می دهد به شرح پژوهش های مرتبط با این حوزه خواهیم پرداخت. در فصل ۳ به تشریح کامل روش پیشنهادی خواهیم پرداخت. در این فصل روش پیشنهادی به شکل بلوکی ارائه شده و سپس در ادامه به شرح کامل هر بلوک و معادلات ریاضی مرتبط با آن خواهیم پرداخت. در فصل ۴ مدلی را بر اساس روش پیشنهادی فصل ۳ طراحی نموده و در خصوص نتایج شبیه سازی بحث خواهیم نمود. همچنین عملکرد و نتایج مدل پیشنهادی با سایر مدل های مشابه مورد مقایسه قرار خواهد گرفت. در فصل ۵ به نتیجه گیری به دست آمده از این پایان نامه خواهیم پرداخت همچنین با بررسی و تجربه چالش های این پژوهش، پیشنهاداتی جهت رفع مشکلات و بهبود الگوریتم در اختیار پژوهشگران قرار خواهد گرفت.

فصل ۲
ادبیات پژوهش

۲ – ۱ – مقدمه

همانگونه که پیش تر بیان شد، فراتفکیک پذیری از حوزه های بسیار پرطرفدار و پرکار خصوصاً در سالیان اخیر بوده است. تا کنون صدها پژوهش در این حوزه انجام شده که می توان آن ها را بر اساس فاکتورهای گوناگون دسته بندی نمود. در این فصل ابتدا به کاربرد های فراتفکیک پذیری خواهیم پرداخت و پژوهش های انجام شده در هر حوزه را معرفی خواهیم نمود. سپس فراتفکیک پذیری را بر اساس نوع الگوریتم، طبقه بندی و بررسی می کنیم. سپس به بررسی اجمالی آخرین مقالات مرتبط با پژوهش خواهیم پرداخت.

۲ – ۲ – طبقه‌بندی فراتفکیک پذیری بر اساس کاربرد

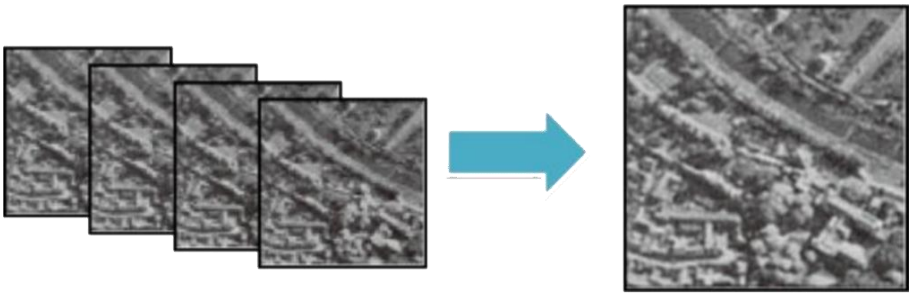
از آنجا که پیاده سازی فراتفکیک پذیری موجب بهبود در عملکرد پردازش تصاویر دیجیتال می گردد، با پیشرفت پردازنده های محاسباتی نرم افزارهای کاربردی بسیاری در ارتباطات دیجیتالی مدرن که اطلاعات را به فرم تصویر و یا ویدیو نگهداری می کنند از تکنیک فراتفکیک پذیری استفاده می کنند. در این قسمت به برخی از مهمترین کاربردهای فراتفکیک پذیری می پردازیم که فراتفکیک پذیری تصویر و ویدیو در آن نقش پررنگی دارند. علاوه بر شرح اجمالی کاربردهای مهم فراتفکیک پذیری، چندی از پژوهش های مرتبط با این کاربردها معرفی و مورد بررسی قرار خواهند گرفت.

۲ – ۲ – ۱ – پردازش تصاویر ماهواره ای

فراتفکیک پذیری نقش بارزی در حوزه تصاویر ماهواره ای دارد. پردازش تصاویر ماهواره ای نیازمند تصحیح[1]، بازیابی[2]، بهبود[3] و استخراج اطلاعات داشته که موجب استفاده متناوب از فراتفکیک پذیری در این حوزه شده است. علاوه بر این فراتفکیک پذیری موجب حذف اعوجاج[4] در تصاویر ماهواره ای شده و اطلاعات جغرافیایی

[1] Rectification
[2] Restoration
[3] Enhancement
[4] Distortion

مورد نیاز را بهبود می دهد [۶] . مقاله [۷] به فراتفکیک پذیری تصاویر ماهواره ای به روش شبکه های همگشتی[1] پرداخته و مقاله [۸] به فراتفکیک پذیری تصاویر ماهواره ای به روش شبکه های مولد تخاصمی فوق متراکم[2] می پردازد.



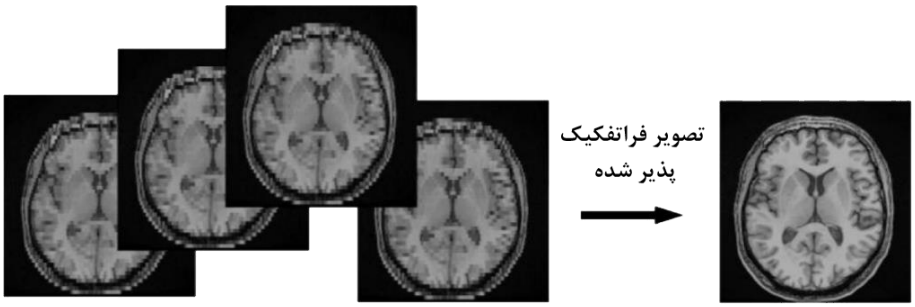
شکل ۱-۲-فراتفکیک پذیری در تصاویر ماهواره ای [۶].

۲ - ۲ - ۲ - پردازش تصاویر پزشکی

نقش فراتفکیک پذیری در پردازش تصاویر پزشکی غیرقابل انکار است. بسیاری از تصاویر پزشکی از وضوح پایین رنج می برند. بهبود تصاویر پزشکی نظیر اسکن CT و MRI نیازمند وضوح بالا و کیفیت قابل قبول می باشند. تصاویر اشعه X کنتراست[3] پایینی داشته و یا تصاویر سونوگرافی حاوی نویز می باشند. علاوه بر این در صورتی که زمان بیشتری برای تصویر برداری مورد نیاز باشد حرکت بیمار موجب مات[4] شدن تصویر می گردد. امروزه بسیاری از ابزارهای پزشکی تصویربرداری پزشکی از کیفیت و سرعت بسیار بالایی برخوردار بوده که به لطف روشهای فراتفکیک پذیری، تصویر خروجی از جزئیات قابل قبولی برخوردار بوده که موجب تشخیص

[1] Convolutional neural networks (CNN)
[2] Ultra-dense GAN
[3] Contrast
[4] Blur

سریع تر بیماری می گردد [۶]. پژوهش [۹] و [۱۰] به فراتفکیک پذیری تصاویر پزشکی به روش شبکه های مولد تخاصمی[1] می پردازند.

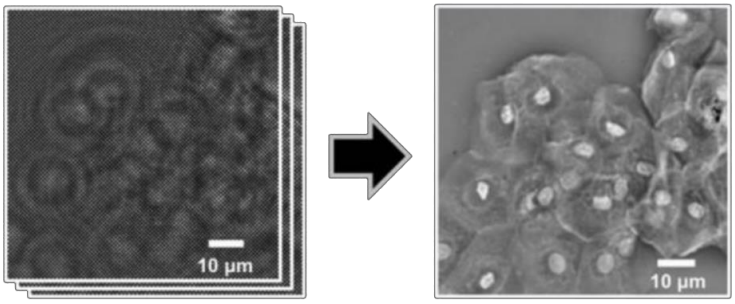


شکل ۲-۲- فراتفکیک پذیری در تصاویر پزشکی [۶].

۲ - ۲ - ۳ - پردازش تصاویر میکروسکوپی

فراتفکیک پذیری نقش مهمی در پردازش تصاویر میکروسکوپی دارد. اخیرا پژوهش های قابل توجهی در خصوص بهبود بصری ساختارهای بیولوژیکال و سلولی تصاویر میکروسکوپی انجام شده که بدون استفاده از متدهای فراتفکیک پذیری میسر نبوده است. به عنوان مثال فراتفکیک پذیری میکروسکوپی فلورسانس یکی از حوزه های قابل توجه در پردازش تصاویر میکروسکوپی می باشد. در سال ۲۰۱۴، جایزه نوبل شیمی مرتبط با حوزه تصاویر میکروسکوپی فلورسانس فراتفکیک پذیر شده بود. [۶] پژوهش های [۱۱] و [۱۲] به بررسی رویکردهای فراتفکیک پذیری مبتنی بر تصاویر میکروسکوپی فلوسانس و سلول های زنده می پردازد.

[1] Generative Adversarial Networks (GAN)

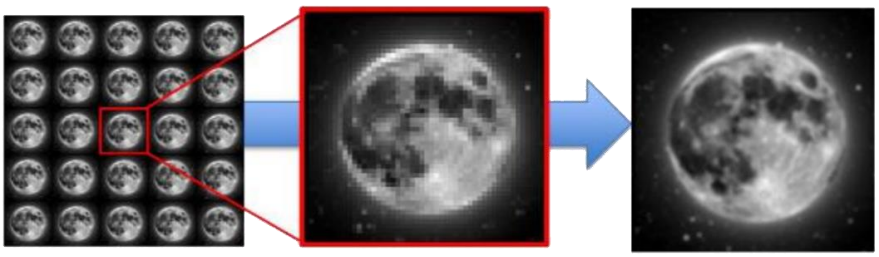


تصاویر خام　　تصویر فراتفکیک پذیر شده

شکل ۳-۲-فراتفکیک پذیری در تصاویر میکروسکوپی [۶].

۲ - ۲ - ۴ - پردازش تصاویر نجومی

تصاویر با وضوح بالا همواره جهت محاسبات بهتر در کاربردهای اخترشناسی و نجومی مورد استفاده بوده است. لذا با تکنیک های فراتفکیک پذیری و با استفاده از ترکیب چندین تصویر مات و نویزی می توان به تصویر با وضوح مطلوب دست یافت [۶]. در پژوهش [۱۳] به بازسازی تصاویر نجومی با استفاده از پیچش نرمالیزه وفقی پرداخته شده است.



تصویر با وضوح پایین　　تصویر فراتفکیک پذیر شده

شکل ۴-۲- فراتفکیک پذیری در تصاویر نجومی [۶].

۲ - ۲ - ۵ - پردازش تصاویر و و ویدیو در صنایع چند رسانه ای

صنایع چندرسانه ای یکی از حوزه های مدرن و بسیار مورد تفاضا در خصوص پردازش تصویر و ویدیو می باشد. با همگانی شدن اینترنت پرسرعت و پیشرفت تلفن های هوشمند اهمیت این حوزه دو چندان شده است. لذا می توان نقش مهم و انکارناپذیری در این حوزه برای تکنیک های فراتفکیک پذیری قائل شد. ارتقا کیفیت تصویر در شبکه های تلویزیونی و اینترنتی، بهبود کیفیت تصویر در مکالمات تصویری آنلاین، افزایش کیفیت تصویر و ویدیو در نرم افزارهای مبتنی بر تصویر و ویدیو بالاخص پیام رسانها، از حوزه های بسیار پرطرفدار و مورد مطالعه بسیاری از پژوهش گران در سالیان اخیر بوده است. علاوه بر این با ظهور تلویزیون های با وضوح بالا نظیر $4K$ و $8K$ و عدم وجود محتوای متناسب با کیفیت با اینگونه نمایشگرها، تکنیک های نرم افزاری نظیر فراتفکیک پذیری برای بازسازی محتوای ویدیویی وضوح بالا بسیار مورد توجه قرار گرفته است. به عنوان مثال [۱۴] جهت فراتفکیک پذیری تصویر راهکاری با کیفیت مناسب و سرعت بالا ارائه داده است که مورد توجه مجموعه‌ی یادگیری ماشین گوگل جهت توسعه ابزار نرم افزاری و سخت افزاری خود قرار گرفته است [۱۵] [۱۶]. پژوهش [۱۷] به فراتفکیک پذیری ویدیو مبتنی بر شبکه های عمیق و با رویکرد تولید فیلترهای بزرگ نمایی پویا می پردازد.

شکل ۵-۲- تصویر بالا به روش درون یابی و پایین به روش فراتفکیک پذیری *SRGAN* بزرگنمایی شده است.

شکل ۲-۵- فراتفکیک پذیری در صنایع چندرسانه ای. تصویر بالا بخشی از یک فیلم را نشان می دهد که به روش درون یابی (شیوه کلاسیک) بزرگنمایی شده است و تصویر پایین بزرگنمایی به روش فراتفکیک پذیری *SRGAN* را نشان می دهد.

۲ - ۲ - ۶ - سایر کاربرد ها

کاربردهای تکنیک های فراتفکیک پذیری صرفاً به موارد ذکر شده محدود نمی شود. پژوهش [۱۸] به تشخیص اشیا[1] با استفاده از فراتفکیک پذیری مبتنی بر شبکه های همگشتی می پردازد. صنایع خودرو و وسائل نقلیه[2] نظیر تشخیص پلاک که در مقاله [۱۹] بدان پرداخته شده، از دیگر کاربردهای فراتفکیک پذیری است. پردازش

[1] Object Detection
[2] Automotive Industry

های بلادرنگ[1] نیز امروزه از اهمیت بسیار زیادی برخودار هستند. مقاله [۲۰] به ارائه فراتفکیک پذیری بلادرنگ جهت استفاده در تلفن ها و گجت های همراه پرداخته است. همچنین کاربردهای نظارتی[2] نظیر پژوهش [۵] نمونه های دیگری از کاربرد فراتفکیک پذیری و اهمیت آن را نشان می دهند.

شکل ۶-۲-کاربرد فراتفکیک پذیری در تشخیص پلاک خودرو. تصویر سمت چپ بزرگنمایی به روش معمولی (درون یابی کلاسیک) و تصویر سمت راست بزرگنمایی به روش فراتفکیک پذیری را نشان می دهد.

۲ – ۳ – طبقه بندی فراتفکیک پذیری بر اساس الگوریتم

الگوریتم های فراتفکیک پذیری در دو دسته بندی کلی خود به دو دسته الگوریتم های مبتنی بر حوزه فرکانس و الگوریتم های مبتنی بر حوزه زمان تقسیم بندی می گردند.

۲ - ۳ - ۱ - حوزه فرکانس

الگوریتم های فراتفکیک پذیری ارائه شده در این گروه ابتدا تصاویر ورودی را به حوزه فرکانس برده و پس از تخمین تصویر با وضوح بالا در این حوزه، تصویر بازسازی شده را به حوزه زمان برمی گردانند. این الگوریتم ها از لحاظ نوع تبدیل مورد استفاده برای انتقال تصویر به حوزه فرکانس می تواند به دو دسته الگوریتم های

[1] Real time processing
[2] Surveillance

مبتنی بر تبدیل فوریه و تبدیل موجک تقسیم بندی گردد. مقاله [۲۱] مثالی از استفاده از روش تبدیل فوریه و مقاله [۲۲] بر مبنای ترکیب روش های تبدیل فوریه و موجک به تفکیک پذیری تصویر می پردازد.

۲ - ۳ - ۲ - حوزه زمان

بسیاری از پژوهش هایی که در ارتباط با فراتفکیک پذیری انجام شده است در حوزه زمان انجام شده است. بسته به مشاهدات در دسترس ورودی، این الگوریتم ها خود می توانند به دو دسته فراتفکیک پذیری چند فریمی[1] و فراتفکیک پذیری تک فریمی[2] تقسیم بندی گردد.

۲ - ۳ - ۲ - ۱ - فراتفکیک پذیری های چند فریمی

این الگوریتم ها که جزو الگوریتم های کلاسیک در این حوزه قرار می گیرند از اولین پژوهش هایی هستند که در حوزه فراتفکیک پذیری به آن پرداخته شده است. در ادامه برخی از این الگوریتمها را بیان می کنیم

الگوریتم های مبتنی بر بک پروجکشن تکراری[3]

در این الگوریتم ها یک حدس احتمالی در مورد تصویر با وضوح بالا ارائه شده و سپس به اصلاح آن در تکرارهای بعدی پرداخته می شود. این کار می تواند بر اساس قرار دادن تصاویر با وضوح پایین در چهارچوب شبکه تصویر وضوح بالا و میانگین گیری از تصاویر وضوح پایین صورت گیرد. در این الگوریتم خطای محاسبه به دست آمده و سعی در به حداقل رساندن آن در تکرارهای بعدی داریم. مقاله های [۲۳]، [۲۴] و [۲۵] در این دسته بندی قرار می گیرند.

[1] Multi-Frame Super Resolution
[2] Single-Frame Super Resolution
[3] Iterative Back Projections

الگوریتم های مبتنی بر فیلترهای وفقی تکراری[1]

الگوریتم های مبتنی بر فیلترهای وفقی تکراری اساساً جهت فراتفکیک پذیری ویدیوهای وضوح پایین جهت تولید ویدیوهای با وضوح بالا توسعه داده شده اند. در این الگوریتم به مسئله به عنوان یک مسئله تخمین وضعیت نگاه می کنیم و لذا جهت حل مسئله از فیلتر کالمن[2] استفاده می کنیم. مقاله های [۲۶]، [۲۷] از جمله پژوهش های مبتنی بر فیلترهای وفقی تکراری هستند.

الگوریتم های مبتنی بر روش مستقیم

اولین الگوریتم های توسعه داده شده فراتفکیک پذیری در این دسته بندی قرار می گیرد. چندین تصویر با وضوح پایین را در اختیار گرفته و یکی از آنها را به عنوان مرجع در نظر می گیریم و باقی تصاویر بر اساس آن در چهارچوب شبکه تصویر قرار می گیرند. سپس تصویر مرجع افزایش مقیاس داده شده و سایر تصاویر در آن ادغام می گردند. تصویر نهایی با وضوح بالا از ادغام تمامی تصاویر در یکدیگر به وجود می آید. در نهایت با استفاده از فیلترهای متوسط گیری[3] به بهبود نتیجه می پردازیم. [۲۳] مقاله های [۲۸] و [۲۹] در زمره این روش ها قرار می گیرند.

۲ - ۳ - ۲ - ۲ - فراتفکیک پذیری تک فریمی

روش های فراتفکیک پذیری تک فریمی از محبوبیت بیشتری نسبت به فراتفکیک پذیری چند فریمی برخوردار است چرا که دسترسی به چندین فریم جهت فراتفکیک پذیری معمولاً کار ساده ای نیست. این روش در زمره روش های مدرن فراتفکیک پذیری قرار می گیرد که خود می تواند به چهاردسته مدل های پیش‌بین، مدل های مبتنی بر لبه، مدل های آماری و مدل های مبتنی بر نمونه تقسیم بندی گردند [۲].

[1] Iterative Adaptive Filtering
[2] Kalman Filter
[3] Median filters

مدل های پیشبین

این دسته بندی از الگوریتم ها تصویر با وضوح بالا را از تصویر با وضوح پایین و از طریق یک فرمول ریاضی از پیش تعریف شده بدون آموزش داده ها به دست می آورند. مدل های مبتنی بر درون یابی نظیر درون یابی دوخطی و درون یابی مکعبی[1] در این دسته قرار می گیرند. مدل های پیش‌بین پیچیدگی محاسباتی کمی داشته و به دلیل اینکه مبتنی بر میانگین وزن دار پیکسل های مجاور عمل می کنند قادر به تولید نواحی صافی در تصویر با وضوح بالا هستند. هرچند مدل های پیشبین در بازسازی لبه ها و بخش های که حاوی فرکانس بالا هستند ناتوان بوده و نتایج مطلوبی را ارائه نمی کنند [۳۰].

مدل های مبتنی بر لبه

لبه ها نقشی مهم در ساختار تصاویر و کیفیت بصری آنها ایفا می کنند. پژوهش های مبتنی بر لبه به یادگیری از طریق استخراج ویژگی های لبه ها جهت بازسازی تصاویر می پردازند. این مدل ها قادر به بازسازی تصاویر با وضوح بالا با لبه های با کیفیت مطلوب و تیز[2] هستند. هرچند این مدل ها تنها قادر به بازیابی جزئیات فرکانس بالایی که مبتنی بر لبه هستند می باشند و در حفظ سایر جزئیات فرکانس بالا نظیر بافت[3] تصویر ناتوان هستند. پژوهش های [۳۱] و [۳۲] از جمله روش های مبتنی بر لبه هستند.

مدل های آماری

جهت پیش‌بینی تصویر با وضوح بالا از تصویر وضوح پایین می توان به استخراج ویژگی های متنوعی از تصویر پرداخت؛ نظیر گرادیان توزیعی که در مقاله [۳۳] بدان پرداخته شده است. همچنین از خصوصیت تنکی[4] گرادیان تصویر نیز می توان برای کاهش محاسبات استفاده نمود. مقاله [۳۴] بدین موضوع پرداخته است .

[1] Bicubic Interpolation
[2] Sharp
[3] Texture
[4] Sparse property

مدل های مبتنی بر نمونه

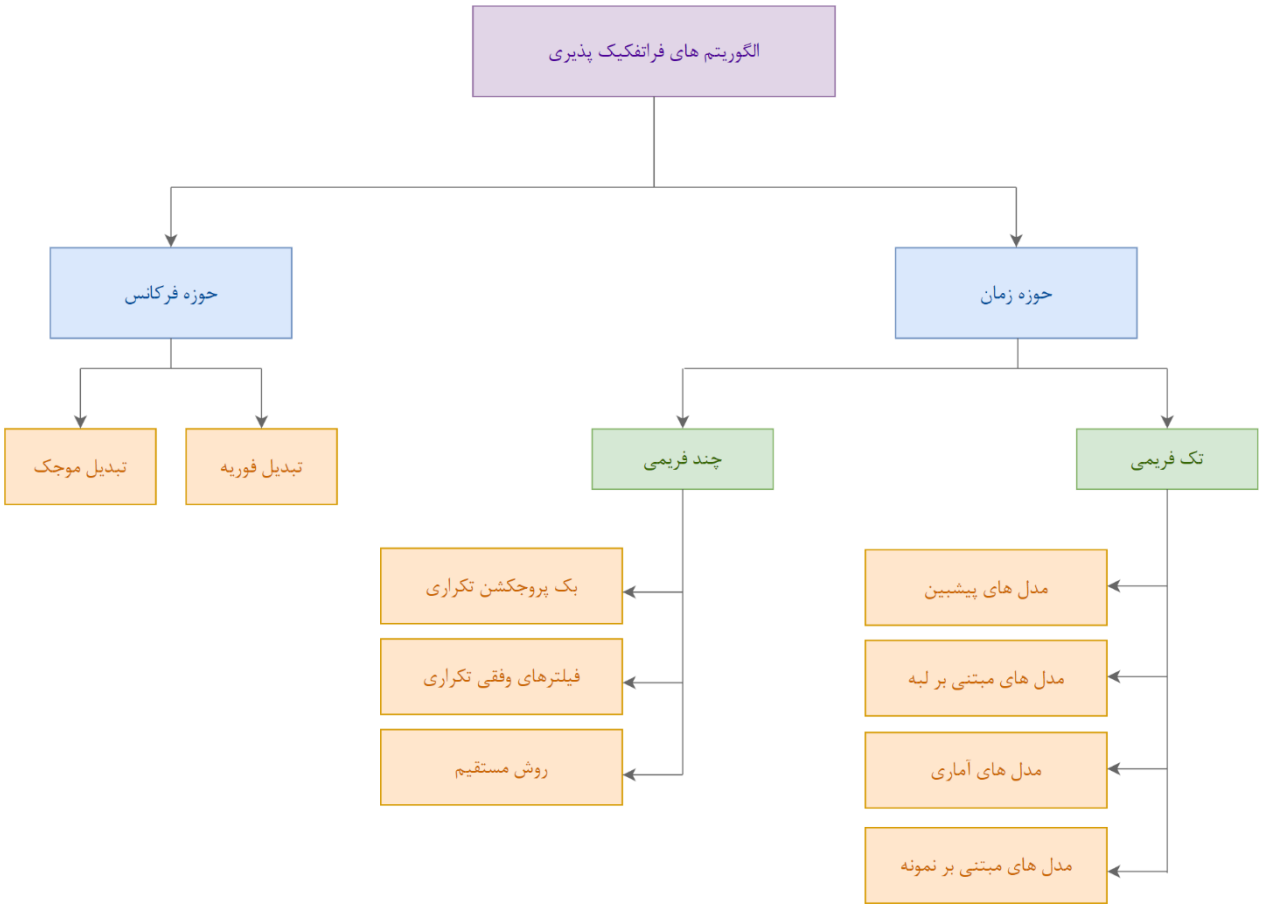
این گونه الگوریتم ها شامل دو مرحله یادگیری[1] و آزمایش[2] بوده که در مرحله یادگیری به ارتباط بین نمونه های با وضوح بالا[3] پرداخته می شود. در این رویکرد ابتدا جفت تصویر وضوح بالا و پایین متناظر به عنوان تصاویر آموزشی به سیستم داده می شود. همچنین تصاویر آموزشی در قالب پنجره‌هایی[4] برش خورده تا بتوان به یادگیری توابع نگاشت پرداخت. اطلاعات استخراج شده در این مرحله به عنوان مبنای گام بعد مورد استفاده قرار خواهد گرفت. انتخاب مجموعه داده در این نوع الگوریتم ها از اهمیت بالایی برخوردار است. هرچند استفاده از مجموعه داده بزرگتر لزوماً به معنای بالابردن دقت نخواهد بود و بالعکس می تواند بار محاسباتی را افزایش داده و موجب اختلال در جستجو نیز گردد. روش های متنوعی جهت یادگیری ماشین می تواند مورد استفاده قرار گیرد. شبکه های عصبی عمیق خصوصاً شبکه های همگشتی از جمله روش های به روز و پراستفاده در فراتفکیک پذیری هستند که از عملکرد بسیار خوبی برخوردار هستند. مقاله های [۲] و [۳۵] به فراتفکیک پذیری تصویر با استفاده از شبکه های همگشتی عمیق پرداخته اند. اخیراً ظهور مدل هایی تحت عنوان شبکه های مولد تخاصمی موجب انقلابی در تولید داده های بسیار واقعی شده است. به کارگیری این الگوریتم در فراتفکیک پذیری که در مقاله [۳۶] بدان پرداخته شده است از جمله روش های پیشتاز در تولید تصاویر با وضوح بالا حتی تا مقیاس ۴ برابر می باشد. روش موجود در این مقاله توانسته به تولید تصاویر فراتفکیک شده ای بپردازد که جزئیات فرکانس بالای تصویر را به خوبی بازسازی نموده و کیفیت بصری بسیار قابل قبولی را ارائه دهد هرچند باید مد نظر داشت که به طور کلی مدل های مبتنی بر نمونه های آموزشی نسبت به سایر روش ها عموماً از پیچیدگی بیشتر و سرعت کمتری برخوردار هستند.

[1] Training
[2] Test
[3] HR
[4] Patch

۲-۳-۳-الگوریتم های ترکیبی

برخی از پژوهش ها جهت بهبود عملکرد از الگوریتم های ترکیبی استفاده می کنند. به عنوان مثال مقاله [۳۷] به بیان رویکردی مبتنی بر الگوریتم لبه یابی و ترکیب آن با مدل های مبتنی بر نمونه پرداخته است.

همچنین لازم به ذکر است که پژوهش های انجام شده صرفا به الگوریتم های مذکور محدود نمی گردند اما عمده ی روش ها را می توان جزو یکی از دسته های مذکور قرار داد. شکل ۲-۷ الگوریتم های مورد بحث را در یک دیاگرام خلاصه نموده است.



شکل ۲-۷-الگوریتم های فراتفکیک پذیری در یک نگاه

۲ – ۴ – فراتفکیک پذیری ویدیو[1]

معرفی پژوهش های انجام شده تا کنون همگی مبتنی بر فراتفکیک پذیری تصویر انجام شد. اما فراتفکیک پذیری را می توان از منظر دیگری در حوزه الگوریتم های مبتنی بر فراتفکیک پذیری ویدیو تقسیم بندی نمود. هرچند فراتفکیک پذیری تصویر و ویدیو دو حوزه ی در هم تنیده در یکدیگر هستند. به گونه ای که می توان فراتفکیک پذیری ویدیو را به نوعی توسعه ای از روش های فراتفکیک پذیری تصویر در نظر گرفت. دلیل اول آنکه فرآیند فراتفکیک پذیری ویدیو در بسیاری از پژوهش ها به ارتقا فریم به فریم ویدیو می پردازد (فراتفکیک پذیری تصویر). در نهایت مجددا با ترکیب فریم های (تصاویر) فراتفکیک پذیرشده، به ویدیو فراتفکیک پذیر شده دست خواهیم یافت. به عبارت دیگر تمام پژوهش های انجام شده در خصوص فراتفکیک پذیری تصویر را می توان به روش بیان شده بر روی ویدیو نیز اعمال کرد. اما از آنجا که فراتفکیک پذیری فریم به فریم ویدیو باعث بروز مشکلاتی نظیر از دست رفتن پیوستگی زمانی[2] و اعوجاج های ساختگی[3] می گردد. علاوه بر این فریم های مجاور در یک ویدیو می توانند حاوی اطلاعات مفیدی جهت بازسازی وضوح بالای یک ویدیو باشند. در این بخش به معرفی مقالاتی خواهیم پرداخت که چالش مذکور را مد نظر قرار داده و علاوه بر فراتفکیک پذیری فریم به فریم ویدیو، به نحوی در تلاش برای کسب اطلاعات حاصل از ترتیب فریم های ویدیو جهت حفظ پیوستگی زمانی در ویدیو های فراتفکیک پذیر شده هستند. متدهای فراتفکیک پذیری ویدیو را می توان در دو دسته بندی کلی به متدهای شامل همترازی[4] و بدون همترازی در نظر گرفت که در ادامه بدان خواهیم پرداخت.

[1] Video Super-Resolution (VSR)
[2] Temporal Consistency
[3] Artifact Distortion
4 Alignment

۲ – ۵ – روش های شامل هم ترازی

در متدهای شامل هم ترازی، فریم های مجاور فریم هدف به طور صریح با استفاده از اطلاعات حاصل از حرکت با فریم کنونی تراز می گردند. روش هایی که در این گروه قرار می گیرند عمدتاً با استفاده از تخمین و جبران حرکت[1] و یا پیچش تغییر شکل پذیر[2] به همترازی فریم ها می پردازند [۳۸].

۲ – ۵ – ۱ – ۱ – روش های مبتنی بر تخمین و جبران حرکت

متدهای مبتنی بر تخمین و جبران حرکت با استفاده از عملیات پیچش[3] فریم های مجاور را با کنونی تراز می کنند. بخش عمده ای از پژوهش ها از شار نوری جهت دریافت اطلاعات ناشی از حرکت استفاده کرده و سعی در محاسبه حرکت بین فریم های مجاور از طریق همبستگی بین این فریم ها در حوزه زمانی دارند [۳۹]. شکل ۲-۸ نحوه انجام این الگوریتم را به تصویر کشیده است.

شکل ۲-۸- مثالی از تخمین و جبران حرکت. در تصویر (د) رنگ موجود در تصویر نمایانگر جهت حرکت و شدت رنگ نمایانگر میزان حرکت بین دو فریم هدف و مجاور را نشان می دهد[۳۸] .

1 Motion Estimation and Motion Compensation (MEMC)
2 Deformable Convolution (DC)
3 Warp

مقاله [۴۰] تحت عنوان فراتفکیک پذیری ویدیو فریم بازگشتی[1]، اطلاعات حاصل از پیوستگی زمانی را با تخمین شار نوری بین فریم قبلی و کنونی محاسبه نموده و پس از ترکیب این اطلاعات با فریم اصلی به فراتفکیک پذیری فریم هدفی می پردازد که حاوی اطلاعات تخمین حرکت قبل نیز می باشد. مدل پیشنهادی این الگوریتم را می توان جزو روش های همترازی و مبتنی بر جبران و تخمین حرکت به حساب آورد.

۲ - ۵ - ۱ - ۲ - روش های مبتنی بر پیچش تغییر شکل پذیر

از سوی دیگر در روش های مبتنی بر شبکه های پیچشی تغییرشکل پذیر به جای استفاده از یک ساختار هندسی ثابت، نگاشت ویژگی فریم هدف را با نگاشت ویژگی فریم ترکیب کرده و آن را تحت لایه های پیچشی اضافه جهت بدست آوردن یک کرنل پیچشی تغییرشکل پذیر به کرنل اصلی اضافه می کند. پژوهش [۴۱] تحت عنوان فراتفکیک پذیری مبتنی بر شبکه های پیچشی تغییرشکل پذیر سه بعدی[2] را می توان جزو این دسته از الگوریتم ها قلمداد نمود.

۲ - ۵ - ۲ - روش های بدون هم ترازی

در مقابل روش های مبتنی بر همت رازی، روش های بدون هم ترازی، فریم های مجاور را با فریم هدف تراز نمی کنند. این روش ها بر اساس استخراج اطلاعات زمانی به دریافت ویژگی های ویدیو می پردازند. این گونه روش ها را می توان به چهار دسته کلی روش های مبتنی بر پیچش دو بعدی[3] ، روش های مبتنی بر پیچش سه بعدی[4] ، شبکه های عصبی پیچشی بازگشتی[5] و شبکه های غیرمحلی[6] تقسیم بندی نمود [۳۸].

1 Frame-Recurrent Video Super Resolution (FRVSR)
2 Deformable 3D Convolution for Video Super-Resolution
3 2D Convolutional Methods (2D Conv)
4 3D Convolutional Methods (3D Conv)
5 Recurrent Convolutional Neural Network (RCNN)
6 Non-local network

۲ - ۵ - ۲ - ۱ - روش های مبتنی بر پیچش دوبعدی

در این روش به جای تراز نمودن فریم های مجاور و هدف جهت تخمین هدف، فریم های ورودی را مستقیماً به یک شبکه پیچشی دو بعدی جهت استخراج ویژگی می دهیم. پژوهش [۴۲] از یک شبکه *GAN* جهت فراتفکیک پذیری ویدیو بر اساس معماری شبکه های پیچشی دوبعدی استفاده کرده است [۳۸].

۲ - ۵ - ۲ - ۲ - روش های مبتنی بر پیچش سه بعدی

روش های مبتنی بر پیچش سه بعدی عملیات استخراج ویژگی را در یک حوزه فضایی-زمانی انجام می دهند. بر خلاف روش های شبکه های پیچشی دوبعدی که عملیات را در حوزه فضایی مدنظر قرار می دهد، در این روش همبستگی زمانی بین فریم ها نیز مد نظر قرار گرفته می شود. پژوهش [۱۷] مبتنی بر شبکه فیلترهای پویا طراحی شده که فیلترها را متناظر با ورودی تولید کرده و در نتیجه ویژگی های متناظری را نتیجه می دهد. ساختار فیلتر افزایش مقیاس پویای طراحی شده که به دریافت اطلاعات فضایی-زمانی ورودی ها می پردازد مبتنی بر شبکه های پیچشی سه بعدی و با اجتناب از تخمین و جبران حرکت طراحی شده است [۳۸].

۲ - ۵ - ۲ - ۳ - روش های مبتنی بر شبکه های عصبی پیچشی بازگشتی

شبکه های عصبی پیچشی بازگشتی عملکرد خوبی را در سایر کاربردهای شبکه های عمیق برجای گذاشته و می تواند در فراتفکیک پذیری ویدیو نیز مورد استفاده قرار گیرد. به طور کلی این شبکه ها جهت مدل سازی اطلاعات فضایی-زمانی فریم ها مناسب بوده و ساختار سبک تری را نسبت به سایر روش های مشابه دارند. هرچند یادگیری این نوع شبکه ها دشوار بوده و بعضاً از مشکلاتی نظیر محوشوندگی گرادیان نیز رنج می برند. پژوهش های [۴۳] تحت عنوان ساخت یک شبکه فضایی-زمانی انتها به انتها جهت فراتفکیک پذیری ویدیو و [۴۴] با عنوان شبکه بازگشتی فضایی-زمانی سریع برای فراتفکیک پذیری ویدیو مبتنی بر شبکه های عصبی پیچشی بازگشتی طراحی شده اند.

۲ - ۵ - ۲ - ۴ - شبکه های غیر محلی

شبکه های غیرمحلی روش دیگری جهت دریافت اطلاعات فضایی-زمانی فریم های ویدیو می باشد. این روش از ایده کلیدی شبکه های عصبی غیرمحلی بهره می برد که نقص شبکه های عصبی پیچشی و بازگشتی در محاسبات محلی را برطرف می کند. پژوهش [۴۵] از بلاک های باقیمانده غیرمحلی برای استخراج ویژگی های فضایی-زمانی یک ویدیو استفاده می کند.

۲ - ۶ - نتیجه گیری

در این فصل به بررسی الگوریتم های مورد استفاده در فراتفکیک پذیری تصویر و ویدیو پرداختیم. نتایج مورد بررسی در این فصل ما را به کلماتی کلیدی نظیر شبکه های عمیق، شبکه های همگشتی، شبکه های مولد تخاصمی و شار نوری به عنوان الگوریتم های پیشرو و مدرن در حل مسئله فراتفکیک پذیری رهنمون می سازد. در فصل آینده به شرح مبانی نظری مورد نیاز برای حل مسئله فراتفکیک پذیری ویدیو می پردازیم.

فصل ۳
مبانی نظری پژوهش

۳ – ۱ – مقدمه

پیش از آنکه به شرح روش پیشنهادی پژوهش حال حاضر که فراتفکیک پذیری ویدیو مبتنی بر شبکه های عمیق می باشد، بپردازیم نیازمند بیان نحوه برخورد با یک مسئله فراتفکیک پذیری خواهیم بود. چطور می توان این مسئله را برای ویدیو بسط داد و در نهایت به بررسی چگونگی حل مسئله فراتفکیک پذیری ویدیو بر اساس شبکه های عمیق خواهیم پرداخت. این فصل به مبانی نظری مورد نیاز در خصوص مسائل مذکور می پردازد و اساس پژوهش حال حاضر را در فصل های آینده تشکیل خواهد داشت.

۳ – ۲ – تعریف مسئله

هدف اصلی فراتفکیک پذیری تک فریمی، تولید یک تصویر با وضوح بالا از طریق یک تصویر با وضوح پایین می باشد. اگر ϑ را عامل مخدوش کننده تصویر[1] بنامیم که کیفیت (میزان نویز) و وضوح (اندازه) تصویر را تعیین می کند، در صورتی که تصویر با وضوح پایین را I^{LR} [2] و تصویر با وضوح بالا را I^{HR} [3] بنامیم. ارتباط ریاضی بین این دو تصویر را می توان به شکل ذیل (۱-۳) بیان نمود.

$$I^{LR} = \vartheta(I^{HR}) \qquad -۱-۳$$

هر تصویر شامل عرض[4] W، ارتفاع[5] H ، کانال های رنگ[6] C و ضریب بزرگ نمایی[7] s می باشد. جدول ۳-۱ معادل انگلیسی اصطلاحات بیان شده جهت استفاده در معادلات ریاضی را بیان می کند. از آنجایی که مسئله ما یک مسئله معکوس بوده، تصویر فراتفکیک پذیر شده از معادله ذیل (۲-۳) به دست می آید:

[1] Degrading factor
[2] Low Resolution Image
[3] High Resolution Image
[4] Width
[5] Height
[6] Color channels
[7] Scaling Factor

$$I^{SR} = \vartheta^{-1}(I^{LR}) \qquad -۲-۳$$

البته تصویر فراتفکیک پذیر شده الزاماً شامل تمام ویژگی های موجود در تصویر وضوح بالا یعنی I^{HR} نخواهد بود. دلیل اصلی این اتفاق وجود تابع ϑ بوده که موجب حذف برخی از اطلاعات تصویر می شود که هم به دلیل افزودن نویز و هم به دلیل کاهش ابعاد تصویر موجب آن می شود. لذا ۳-۲ یک تابع یک به یک را نتیجه نمی دهد. به عبارت دیگر فراتفکیک پذیری یک مسئله معکوس بوده و جواب یکتایی نخواهد داشت. به همین دلیل می توان آن را جزو مسائل بدطرح[1] قلمداد نمود.

از آنجایی که فراتفکیک پذیری ویدیو تعمیمی از فراتفکیک پذیری تصویر می باشد، می توان ۳-۱ و ۳-۲ را به فراتفکیک پذیری ویدیو تعمیم داد. با فرض وجود N فریم در یک ویدیو I_t^{HR} فریم با وضوح بالای یک ویدیو در زمان t و I_t^{LR} فریم با وضوح پایین متناظر آن می باشد. در نتیجه خواهیم داشت:

$$\sum_{t=1}^{N} I_t^{LR} = \sum_{t=1}^{N} \vartheta_t(I_t^{HR}) \qquad -۳-۳$$

در نتیجه برای به دست آوردن ویدیوی فراتفکیک پذیر شده نیاز به حل مسئله ذیل خواهیم داشت:

$$\sum_{t=1}^{N} I_t^{SR} = \sum_{t=1}^{N} \vartheta_t^{-1}(I_t^{LR}) \qquad -۴-۳$$

۳ ـ ۳ ـ حل مسئله فراتفکیک پذیری به روش شبکه های عمیق

همواره در حل مسئله فراتفکیک پذیری به روش شبکه های عمیق چهار پارامتر ذیل مورد بحث قرار می گیرد. روش افزایش مقیاس[2]، معماری شبکه[3]، توابع هزینه[4] و معیارهای عملکرد[5] مدل.

[1] Il-posed
[2] Upscaling
[3] Network Architecture
[4] Loss Functions
[5] Performance Metrics

پژوهشگران با بهبود و توسعه رویکرد خود و ارائه ی روش های نوین در هر یک از پارامترهای فوق که چهارچوب حل مسئله فراتفکیک پذیری به روش شبکه های عمیق را شکل می دهد سعی در بهبود عملکرد فراتفکیک پذیری و کیفیت خروجی دارند. پژوهش حال حاضر از این منطق مستثنی نیست. هر چند قبل از بیان رویکرد پیشنهادی، بهتر است به شرح پارامترهای مذکور پرداخته شود.

جدول ۱-۳-اصطلاحات و نمادهای مورد استفاده در معادلات ریاضی

اصطلاح	نماد
تصویر وضوح بالا	I^{HR}
تصویر وضوح پایین	I^{LR}
تصویر فراتفکیک پذیر شده	I^{SR}
فریم فراتفکیک پذیر شده در زمان t	I_t
فریم فراتفکیک پذیر شده ماقبل t	I_{t-1}
عرض تصویر	W
ارتفاع تصویر	H
کانال های رنگ تصویر	C
ضریب بزرگ نمایی	s

۳ - ۳ - ۱ - روش های افزایش مقیاس

افزایش مقیاس تصویر ورودی جهت به دست آوردن خروجی وضوح بالا یکی از مؤلفه های اصلی در فراتفکیک پذیری می باشد. بعضی از روش ها به نگاشت ویژگی تصاویر وضوح پایین برای تولید تصویر وضوح بالای متناظر می پردازند. هرچند روش های قدیمی تر قادر به نگاشت ویژگی نیستند و به جای آن از خود تصویر به عنوان ورودی استفاده می کنند. در این بخش به تعدادی از مهمترین روش های افزایش مقیاس در فراتفکیک پذیری اشاره می کنیم.

۳ - ۳ - ۱ - ۱ - درون یابی دومکعبی

این روش افزایش مقیاس یک روش کلاسیک جهت افزایش مقیاس تصویر داده شده با اعمال درون یابی مبتنی بر پیکسل های اصلی تصویر می باشد. با فرض داشتن تصویر ورودی، این تکنیک ابتدا پیکسل های تصویر اصلی را روی یک صفحه تصویر بزرگ شده قرار داده و سپس پیکسل های بدون مقدار را با استفاده از درون یابی بین ۱۶ پیکسل مجاور (4×4) خود انجام می دهد.

این تکنیک در برخی از مدل های فراتفکیک پذیری به عنوان بخشی از مدل به کار گرفته می شود. به عنوان مثال در شبکه های همگشتی، شبکه به نگاشت تصویر درون یابی شده به متناظر فراتفکیک پذیر شده خود و با استفاده از یادگیری جزئیات از دست رفته می پردازد. این روش از سرعت بسیار بالایی برخوردار است، هرچند بدون ارائه ویژگی بیشتر صرفاً پارامترهای تصویر را در روش های مبتنی بر یادگیری افزایش می دهد [۳۹].

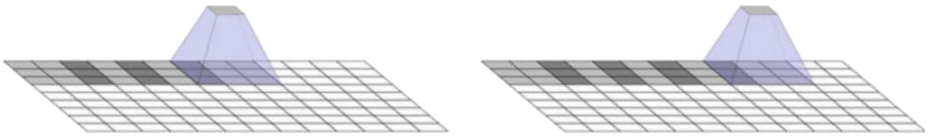
۳ - ۳ - ۱ - ۲ - لایه همگشتی ترانهاده[1]

این لایه که به لایه واهمگشتی[2] نیز شناخته می شود دقیقاً عکس یک لایه همگشتی عمل می کند. به عبارت دیگر ویژگی های تصویر به عنوان ورودی داده شده و لایه همگشتی سعی در پیش بینی تصویر ورودی که مرتبط با ویژگی ها بودند، می کند.

جهت جلوگیری از افزایش پارامترهای شبکه که به دلیل افزایش مقیاس تصویر قبل از ورود آن به شبکه انجام می شود، ابتدا نگاشت های ویژگی می تواند از فضای تصویر وضوح پایین انجام شده و سپس یک لایه همگشتی ترانهاده در انتهای شبکه به بازسازی تصویر وضوح بالا از طریق ویژگی های استخراج شده اقدام نماید. بدین صورت بلوک های لایه های همگشتی نیازی به پردازش تصاویر درون یابی شده ندارند و موجب کاهش پیچیدگی محاسباتی شبکه می شوند. لازم به ذکر است که لایه همگشتی ترانهاده خود یک لایه مبتنی بر

[1] Transpose Convolution Layer
[2] Deconvolutional

یادگیری بوده در حالی که درون یابی یک الگوریتم ساده و بی ارتباط با ورودی می باشد. اگرچه که رویکرد مذکور منجر به عملکرد بهتر شبکه خواهد شد اما موجب اثر تصنعی شطرنجی[1] در تصویر خروجی می گردد. دلیل این امر این است که ابعاد پنجره ضریب صحیحی از گام حرکت[2] نیست. شکل ۳-۱ به خوبی این مسئله را نشان داده است. این گونه همپوشانی ها موجب اثر تصنعی شطرنجی می شود که در شکل ۳-۲ بدان پرداخته شده است [۳۹].



شکل ۱-۳- همپوشانی فیلترهای ۳x۳ که قابل تقسیم صحیح بر گام حرکت ۲ نیست [۳۹].

شکل ۲-۳- مثالی از وجود اثر تصنعی شطرنجی در تصویر [۳۹].

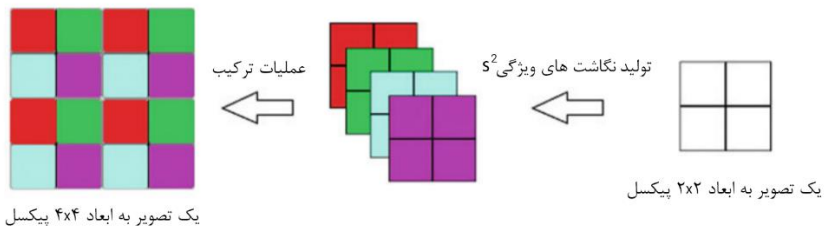
[1] Checkboard Artifacts
[2] Stride

۳ - ۳ - ۱ - ۳ - لایه همگشتی *sub-pixel*

پژوهش [۴۶] جهت غلبه بر مشکلی که در نتیجه لایه همگشتی ترانهاده رخ می دهد، به بیان رویکردی متفاوت پرداخته است. در این رویکرد جهت تولید یک تصویر با وضوح بالا، پیکسل های تصویر را تحت نگاشت های ویژگی متنوعی ترکیب[1] می کنیم. این پژوهش راهکار گفته شده را تحت عنوان لایه همگشتی *sub-pixel* طرح می کند.

بدین منظور اگر یک تصویر وضوح پایین را به شکل $H \times W \times C$ با ضریب افزایش مقیاس s بیان کنیم، لایه همگشتی *sub-pixel* به تعداد s^2 نگاشت ویژگی با لایه همگشتی قراردادی تولید می کند. این بخش ها به اندازه $H \times W \times s^2C$ با هم ترکیب شده که بتواند یک تصویر با وضوح بالا به ابعاد $sH \times sW \times C$ تولید نماید. شکل ۳-۳ این عملیات را به خوبی نشان می دهد.

لایه همگشتی *sub-pixel* به دلیل عدم وجود مشکل اثر تصنعی شطرنجی نسبت به لایه همگشتی ترانهاده از محبوبت بالاتری برخوردار است و در پژوهش های بیشتری مورد استفاده قرار گرفته است [۳۹].



شکل ۳-۳- اعمال ترکیب کننده پیکسلی با ضریب افزایش مقیاس ۲ (*s=2*) [۳۹].

[1] reshuffle

۳ - ۳ - ۲ - معماری شبکه

معماری شبکه نقش بسیار مهمی در فراتفکیک پذیری به روش شبکه های عمیق دارد. در این بخش به پرکاربردترین معماری های شبکه های مورد استفاده در این حوزه می پردازیم.

۳ - ۳ - ۲ - ۱ - یادگیری مبتنی بر باقیمانده[1]

آموزش مبتنی بر باقیمانده اولین بار توسط هی و همکاران[2] در مقاله [۴۷] مطرح شد و بعدها به شکل گسترده ای در معماری شبکه های یادگیری عمیق مورد استفاده قرار گرفت. این معماری نسبت به سایر رویکردهای هم نوع خود از سرعت بالاتری برخوردار است. در یک مسئله فراتفکیک پذیری، یادگیری مبتنی بر باقیمانده ابتدا سعی در یادگیری جزئیات فرکانس بالای تصویر که با جزئیات فرکانس پایین به دست آمده از تصویر وضوح پایین ترکیب شده است دارد. جزئیات فرکانس پایین می تواند از زنجیره ای از لایه های همگشتی به دست آید. لذا باقیمانده را می توان به شکل ذیل (۳-۵) به شکل ریاضی نوشت:

$$Residual = \{I_{HR} - \chi(w, h, c)\} \qquad -۵-۳$$

که χ نمایانگر ویژگی های استخراج شده از لایه های همگشتی می باشد. با استفاده از یادگیری مبتنی بر باقیمانده، شبکه از بار وظایف یادگیری اطلاعات اضافی که قبلاً در تصویر با وضوح پایین موجود بود رهایی می یابد. یادگیری مبتنی بر باقیمانده به طور کلی می تواند تحت دو دسته بندی ذیل تعریف شوند:

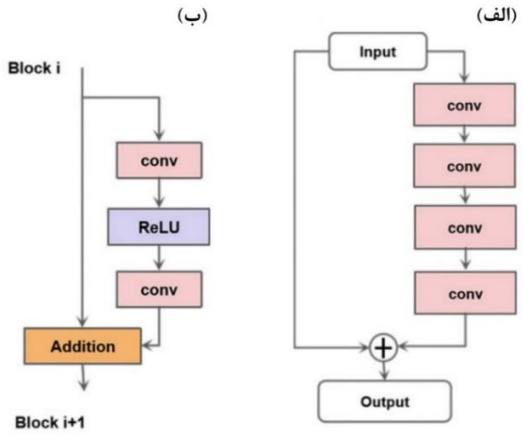
[1] Residual
[2] He et al.

یادگیری مبتنی برباقیمانده سراسری (*GRL*)[1]

متد *GRL* تصویر ورودی را تا انتهای شبکه با معرفی یک اتصال پرش[2] از ورودی به لایه انتهایی شبکه حمل می کند. همانطور که در شکل ۳-۴ مشخص است به دلیل اینکه تصویر ورودی و تصویر خروجی به شدت به یکدیگر همبسته هستند، با فراهم کردن یک راه میانبر می توان بار حمل اطلاعات اضافی را کاهش داد [۳۹].

یادگیری مبتنی بر باقیمانده محلی (*LRL*)[3]

متد *LRL* نگاشت های ویژگی را به جای کل لایه ها، با معرفی اتصالات پرش بین لایه های مختلف همگشتی انجام می دهد. این رویکرد جهت مقابله با حذف جزئیات که در لایه های عمیق همگشتی روی میدهد با استفاده از اضافه کردن جزئیات از لایه های قبلی به لایه دیگر انجام می شود. رویکرد گفته شده یک موضوع حیاتی و بسیار مهم در شبکه های یادگیری بسیار عمیق می باشد. چرا که می تواند مشکلاتی ناشی از همگرایی کند را برطرف نماید [۳۹].



شکل ۴-۳- (الف)یادگیری مبتنی بر باقیمانده سراسری (GRL) ب)-یادگیری مبتنی بر باقیمانده محلی (LRL) [۳۹].

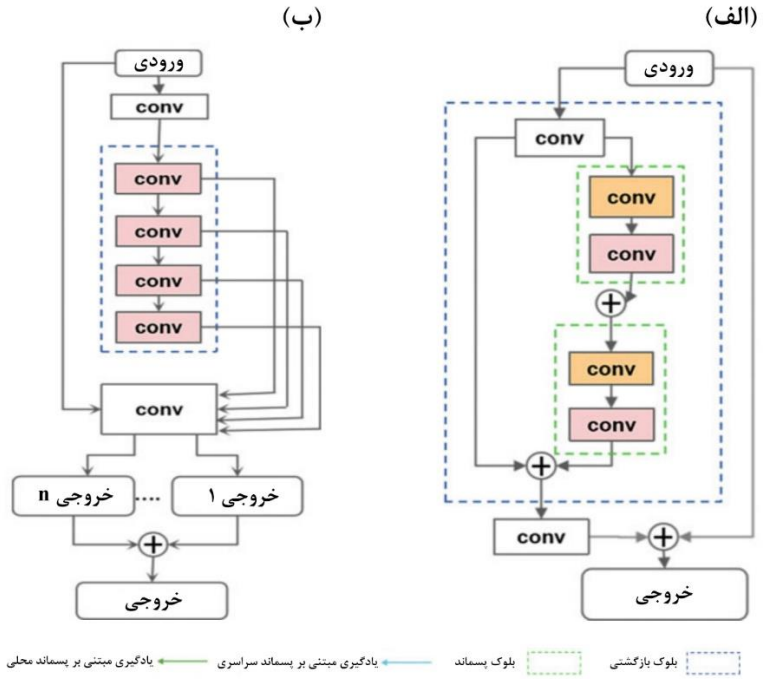
[1] Global Residual Learning (GRL)
[2] Skip Connection
[3] Local Residual Learning (LRL)

۳ - ۳ - ۲ - ۲ - معماری بازگشتی

در حالی که تمامی رویکردهایی که پیشتر در خصوص معماری شبکه های عمیق گفته شد جزو تکنیک های به روز و دارای عملکرد مناسب می باشند، همچنان از دیدگاه مصرف حافظه و پردازش بهینه نیستند. دلیل این امر اینست که هرچه شبکه ها عمیق تر گردد اندازه مدل با افزایش تعداد پارامترها گسترش می یابد.

در حالی که عمیق تر نمودن شبکه ها می تواند موجب بهبود عملکرد مدل گردد اما پردازش های سنگین و مصرف حافظه زیاد عملاً استفاده از این سیستم ها را خصوصاً در تلفن ها و گجت های همراه ناممکن و یا محدود می کند. جهت رفع این مشکل کیم و سایر همکاران[1] در مقاله [۴۸] روشی به نام *DRCN* ارائه دادند که متمرکز بر بلوک های بازگشتی می باشد. همانطور که در شکل ۳-۴ نشان داده شده است هر بلوک بازگشتی شامل زنجیره ای از لایه های همگشتی بوده که همگی از وزن عصبی[2] یکسانی برخوردار هستند. اشتراک این ضرایب در کل لایه ها موجب کاهش پارامترهای شبکه می گردد. پژوهش [۴۹] روشی به نام *DRRN* ارائه داد که به بهبود عملکرد تکنیک پیشنهادی با استفاده از معرفی چندین ضریب اشتراک و تلفیق آن با معماری *LRL* می پردازد. در شکل ۳-۴ ضرایب اشتراکی نشان داده شده است لایه های همگشتی یک رنگ شامل وزن یکسان به اشتراک گذاشته هستند.

[1] Kim et al.
[2] Neural Weights

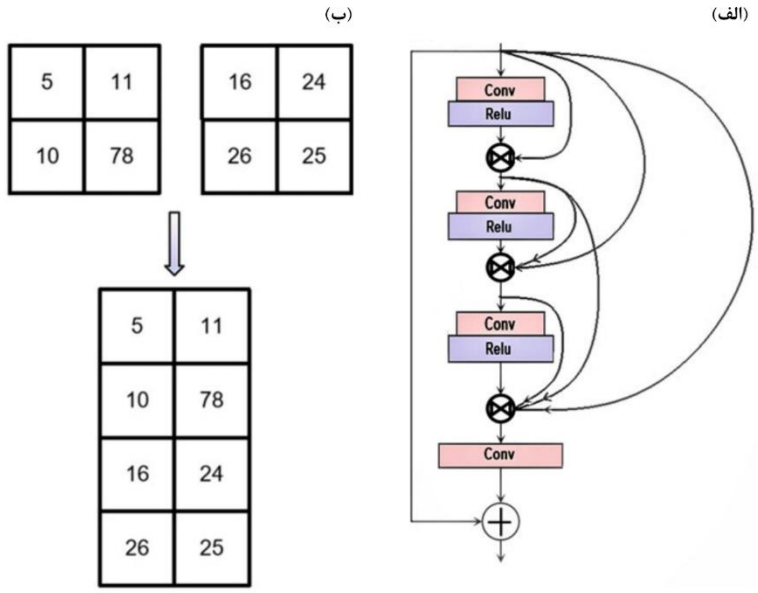


شکل ۵-۳- (الف) - شبکه بازگشتی دو بلوکه در DRRN (ب) – بلوک بازگشتی خطی در DRCN [۳۹].

۳ - ۳ - ۲ - ۳ - شبکه متراکم

اتصالات متراکم اولین بار توسط هوانگ و همکاران[1] در پژوهشی به نام *DenseNet* معرفی شد [۵۰]. پس از آن تانگ و همکاران[2] با بهبود این شبکه عملکرد فوق العاده ای را در بین جدیدترین روش های موجود از خود برجای گذاشتند[۵۱] . شبکه متراکم ممکن است در نگاه اول چیزی جز تکرار بلوک های باقیمانده با اتصالات پرش بیشتر نباشد. هرچند اساساً این دو شبکه دارای دو تفاوت مهم هستند. اول اینکه شبکه متراکم شامل اتصال پرش از یک لایه همگشتی به هر لایه بعدی خود در بلوک می باشد و نگاشت ویژگی ها را به هم الحاق[3] می کند. شکل ۳-۵ این عملیات را به خوبی نشان می دهد.

[1] Huang et al.
[2] Tong et al.
[3] Concatenate



شکل ۶-۳- (الف) - نمایش اتصالات در شبکه متراکم (ب) – عملیات الحاق ویژگی ها [۳۹].

دوم اینکه رشد شبکه به اندازه بلوک های باقیمانده به پیچیدگی شبکه اضافه نمی کند. اتصالات پرش شبکه را به استفاده مجدد از نگاشت ویژگی های قبلی تشویق می کند. این رویکرد موجب می شود که یادگیری شبکه در خصوص ارتباط بین پیکسل ها بهبود پیدا کند.

۳ - ۳ - ۳ - توابع هزینه

توابع هزینه نقشی اساسی در شبکه های عمیق و بیان نحوه بهینه سازی شبکه دارند. انتخاب توابع هزینه متفاوت ما را به مقادیر متفاوتی در خصوص ضرایب شبکه رهنمون می سازد. لذا انتخاب تابع هزینه نقشی کلیدی در مسائل یادگیری ماشین دارند. توابع هزینه متفاوتی در طی سالیان گذشته و طی پژوهش های مختلف مورد استفاده قرار گرفتند که ما در اینجا به مهمترین توابع هزینه مورد استفاده می پردازیم:

۳ - ۳ - ۳ - ۱ - تابع هزینه پیکسلی[1] (*MSE*)

این تابع هزینه که به عنوان میانگین مربع خطا (*MSE*) نیز نامیده می شود تابع هزینه را بر مبنای تفاوت پیکسل به پیکسل تصویر خروجی نسبت به تصویر اصلی محاسبه می کند. تابع هزینه پیکسلی به طور گسترده ای در مدل های مبتنی بر آموزش مورد استفاده قرار می گیرد. *PSNR*[2] که در بخش ۳-۳-۴-۲ به طور مفصل در مورد آن صحبت خواهد شد به عنوان یک معیار عملکرد شبکه معرفی می گردد و به طور معکوس با تابع هزینه *MSE* در ارتباط است. بهینه سازی *MSE* افزایش *PSNR* را در فراتفکیک پذیری تصویر نتیجه می دهد. هرچند استفاده از این معیار جهت مقایسه بین خروجی و تصویر اصلی مناسب به نظر می رسد اما در تولید جزئیات و بافت تصویر ناتوان به نظر می رسد. علت این امر این است که تابع هزینه پیکسلی همانند یک فیلتر میانگین گیری عمل نموده و موجب از دست رفتن جزئیات فرکانس بالای تصویر می گردد.

۳ - ۳ - ۳ - ۲ - تابع هزینه ادراکی[3]

تابع هزینه ادراکی اولین بار توسط جانسون و همکاران[4] در پژوهش [۵۲] مطرح شد با استفاده از یک مدل از پیش آموزش داده شده به محاسبه تابع هزینه جهت آموزش شبکه می پردازد. نحوه عملکرد این تابع هزینه بدین صورت است که ابتدا تصویر اصلی و تصویر فراتفکیک پذیر شده به یک مدل از پیش آموزش داده شده خورانده می شود. سپس لایه های فعال ساز از یک مدل از پیش آموزش داده شده تحت یک ماتریس استخراج می شود. سپس ماتریس به دست آمده از فعال سازهای تصویر اصلی و فراتفکیک پذیر شده جهت محاسبه تابع هزینه محاسبه می شوند.

---

[1] Pixel Loss
[2] Peak signal-to-noise ratio
[3] Perceptual loss
[4] Johnson et al.

ایده اصلی این روش بهره گیری از مدل های از پیش آموزش داده شده معینی بوده که جهت یادگیری کیفیت معنایی و بصری تصاویر در پردازش تصویر مورد استفاده قرار می گیرند. به دلیل همین ویژگی این مدل ها، لایه های عمیق تری می تواند جهت تعریف کیفیت ادراکی تصویر مورد استفاده قرار گیرند. چرا که این مدل ها برای استخراج ویژگی های خاصی آموزش داده شده اند. یک مثال از مدل های پراستفاده مدل *VGG19* می باشد. به طور خاص تر لایه دوم از بلوک پنجم این مدل به طور ویژه ای جهت محاسبات توابع هزینه ادراکی مورد استفاده قرار می گیرد. تابع هزینه ادراکی نسبت به تابع هزینه پیکسلی مناسب تر می باشد چرا که ویژگی های بصری تصویر را به جای مقایسه پیکسل به پیکسل بهبود می دهد [۳۹].

۳ - ۳ - ۳ - ۳ - تابع هزینه تخاصمی[1]

با ورود شبکه های مولد تخاصمی به حوزه فراتفکیک پذیری توسط لدیگ و همکاران در پژوهش [۳۶] تابع هزینه جدیدی به نام تابع هزینه تخاصمی پیشنهاد شد. بلوک متمایزگر پیشنهادی توسط همان مجموعه داده موجود در سمت بلوک مولد آموزش داده شده که قادر به تفکیک تصویر فراتفکیک پذیر شده و اصلی باشد. تابع هزینه تخاصمی شامل مجموع هزینه مولد و متمایزگر می باشد.

با معرفی هزینه تخاصمی، بلوک مولد سعی در فریب متمایزگر با تولید جزئیات و بافت با کیفیت تر در تصاویر می باشد. همزمان بلوک متمایزگر به بهبود خود در زمینه تشخیص تصاویر وضوح بالای اصلی و فراتفکیک پذیر شده می پردازد. این موضوع موجب افزایش عملکرد کلی سیستم در بهبود کیفیت بصری تصویر می گردد [۳۹].

[1] Adversarial loss

۳ - ۳ - ۴ - معیارهای عملکرد مدل

۳ - ۳ - ۴ - ۱ - معیار *SSIM* [۱]

معیار شباهت ساختاری (*SSIM*) کیفیت بصری تصویر فراتفکیک پذیر شده را نسبت به تصویر اصلی از دیدگاه روشنایی و تباین بررسی می کند. معیار *SSIM* وابستگی پیکسل های از لحاظ موقعیت مکانی نزدیک به هم را بررسی کرده تا بتواند بر این اساس کیفیت بصری تصویر فراتفکیک پذیری را مورد سنجش قرار دهد.

چشم انسان به نور موجود در تصویر حساسیت بیشتری نشان می دهد و معیار *SSIM* روشنایی را به عنوان یکی از فاکتورهای خود جهت سنجش کیفیت تصویر مد نظر قرار می دهد. به همین دلیل معیار *SSIM* می تواند معیار بهتری در خصوص سنجش کیفیت تصویر نسبت به *PSNR* باشد.

۳ - ۳ - ۴ - ۲ - معیار *PSNR*

نسبت قله سیگنال به نویز متناوباً جهت سنجش کیفیت تصاویر تولیدی در پردازش تصویر مورد استفاده قرار می گیرد. هرچند همانطور که بیان شد این معیار کیفیت تصویر را نمی تواند تضمین کند، چرا که بررسی پیکسل های متناظر الزماً تامین کننده کیفیت تصویر نیست و همواره یک تصویر با *PSNR* بالا از کیفیت بالاتری در صحنه های واقعی برخوردار نیست. معیار *PSNR* علی رغم مشکلات گفته شده همچنان در بسیاری از پژوهش ها به عنوان یک معیار کیفیت تصویر استفاده می شود [۳۹].

[۱] Structural Similarity Index Measure (SSIM)

۳ – ۴ – بررسی دو مدل مبتنی بر شبکه های عمیق

در این پژوهش از دو مدل شبکه های مولد تخاصمی و شار نوری در روش پیشنهادی استفاده شده است. لذا لازم می دانیم که قبل از استفاده از این دو شبکه در مدل پیشنهادی، به شرح مختصری از این دو مدل بپردازیم.

۳ - ۴ - ۱ - شبکه های مولد تخاصمی (GAN)

شبکه های مولد تخاصمی را به می توان یک نوع طبقه بندی از یادگیری ماشین در نظر گرفت. این رویکرد که اولین بار توسط ایان گودفلو[1] و همکارانش در سال ۲۰۱۴ مطرح شد مبتنی بر آموزش دو شبکه عصبی در رقابت با یکدیگر در یک بازی جمع صفر هستند. این مدل به آموزش همزمان دو بلوک به نام های مدل مولد (G)[2] و مدل متمایزگر (D)[3] می پردازد. به طور کلی مدل مولد G وظیفه تولید داده های ساختگی را برعهده دارد و مدل متمایزگر D احتمال اینکه نمونه از مجموعه داده آموزشی انتخاب شده و یا توسط مدل G ساخته شده باشد را تخمین می زند. روند آموزش مدل G بر مبنای حداکثر نمودن اشتباه مدل D اتفاق می افتد. در این چهارچوب، مدل مولد را می توان یک تیم جعل کننده اسکناس در نظر گرفت که سعی در تولید اسکناس های تقلبی بدون آنکه قابل شناسایی باشند را دارند. مدل متمایزگر را می توان یک تیم پلیس در نظر گرفت که سعی در تشخیص اسکناس جعلی از واقعی دارد. رقابت بین این دو تیم موجب بهبود تجربه و رویکردهای هر دو تیم (مدل) خواهد شد. این رقابت تا جایی اتفاق می افتد که تیم جعل کننده (مدل مولد) قادر به تولید اسکناس هایی (داده هایی) باشد که پلیس (مدل متمایزگر) قادر به تفکیک آن از اسکناس واقعی نباشد.

از همین ایده می توان در حوزه فراتفکیک پذیری و تولید تصاویر با وضوح بالا استفاده نمود. در این رویکرد مدل G قرار است به نحوی آموزش ببیند که به تولید تصاویر فراتفکیک شده ای بپردازد که موجب فریب مدل متمایزگر D در تشخیص تصویر فراتفکیک پذیر شده از اصلی می گردد. نتیجه به تصاویر فراتفکیک پذیر شده

[1] Ian Goodfellow
[2] Generator
[3] Discriminator

طبیعی و بسیار واقعی ختم می شود که به سختی توسط یک ناظر انسانی نسبت به تصویر اصلی وضوح بالا قابل تفکیک می باشد.

۳ - ۴ - ۲ - شار نوری

شار نوری یکی از مسائل مورد طرح در بینایی ماشین[۱] می باشد. مفهوم شار نوری در واقع به توزیع حرکت ظاهری الگوی روشنایی در تصویر گفته می شود. به بیان دیگر تغییرات حرکتی آشکار در شدت روشنایی تصویر را می توان برابر با مفهوم شار نوری در نظر گرفت. شکل ۳-۷ شار نوری موجود در یک فریم از ویدیو را نمایش داده است.

شکل ۷-۳- نمایش شار نوری در یک فریم از ویدیو

[۱] Computer Vision

مطالعات انجام شده روی شار نوری برای چندین دهه مبتنی بر رویکردی بود که توسط هورن و شانک[1] در پژوهش [۵۳] انجام شد. پژوهش مذکور شار نوری را به عنوان یک مسئله کمینه نمودن انرژی در نظر گرفته که تمایل به پیوستگی روشنایی بین پیکسل های مرتبط را دارد. پژوهش های آینده سعی در بهبود این مدل داشتند. هرچند که مدل های ارائه شده تماماً دچار مشکلات و چالش هایی از جمله جابه جایی های مکانی محسوس، تغییرات بریده روشنایی و و چالش های انسداد[2] بودند. در این اثنا، با ظهور شبکه های پیچشی عمیق و نتایج قابل توجه آن در مسائل مرتبط با طبقه بندی تصاویر، پای این مدل نیز به سایر مسائل بینایی ماشین از جمله شار نوری باز شد. لذا از شار نوری مبتنی بر شبکه های پیچشی عمیق به عنوان یک استخراج کننده ویژگی پیشرفته در تصاویر یاد می شود که نتایج بسیار مطلوب تری را نسبت به سایر رویکردهای کلاسیک از خود بر جای می گذارد. در مدل جدید شبکه های پیچشی عمیق این امکان را به ما می دهد که بتوانیم به جای استفاده از ویژگی های دست ساز نظیر گرادیان تصویر به یادگیری بردار ویژگی های استخراج شده بعد بالا[3] به ازای هر پیکسل بپردازیم که نتایج قابل ملاحظه و مقاوم تری را با تغییر شرایط و ظاهر تصویر نسبت به روش های کلاسیک شار نوری از خود بر جای می گذارد [۵۴].

در فصل بعد به معرفی و شرح روش پیشنهادی خواهیم پرداخت.

[1] Horn and Schunck

[2] Occlusion

[3] High Dimensional

فصل ۴
روش پیشنهادی

۴ – ۱ – مقدمه

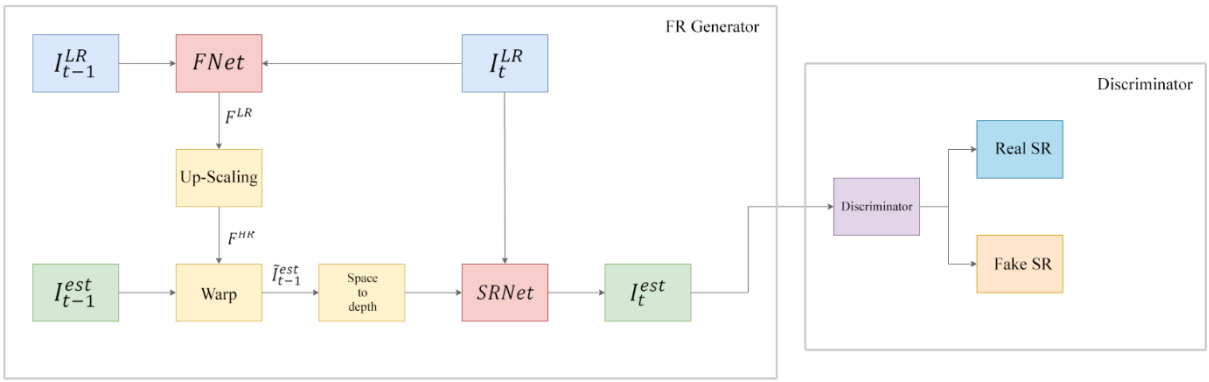
پژوهش های اخیر در فراتفکیک پذیری (همانطور که در فصل ۲ به تفصیل به آن پرداختیم) نشان داده اند که رویکردهای مبتنی بر آموزش به خصوص شبکه های عمیق نقش بسیار موثری در افزایش دقت و کیفیت فراتفکیک پذیری ایفا می کنند. در روش پیشنهادی سعی بر طراحی مدلی داریم که بتواند به بهترین نحو از شبکه های عمیق جهت بازیابی فریم به فریم (تصویر) استفاده نماید. هر چند فارق از فراتفکیک پذیری فریم به فریم (تصویر) به دنبال راه حلی جهت حفظ ساختار ویدیو در این بازسازی خواهیم بود.

اصلی ترین نکته ای که رویکردهای مبتنی بر فراتفکیک پذیری ویدیو را از فراتفکیک پذیری تصویر جدا می کند حفظ ساختار پیوستگی زمانی آن است. توضیح بیشتر آنکه زمانی که یک ویدیو را به مجموعه ای از فریم ها تقسیم می کنیم، نه تنها هر فریم حاوی اطلاعات ساختاری منحصر به فرد خود است بلکه شامل اطلاعات اضافه تری بوده که از ترتیب خاص قرارگیری فریم ها در کنار هم جهت تولید ویدیو به دست می آید. بازیابی اینگونه اطلاعات اضافی موجب نرم تر و به طور کلی مطلوبیت بیشتر ویدیوی فراتفکیک پذیر شده می گردد و موجب شده است که فراتفکیک پذیری ویدیو به مسئله جالب تر و در عین حال چالش برانگیز تری نسبت به فراتفکیک پذیری تصویر قلمداد شود. روش پیشنهادی به رویکردی جهت دریافت این اطلاعات اضافی نیز می پردازد.

در این فصل به شرح چهارچوب کلی روش پیشنهادی و الگوریتم مورد استفاده پرداخته و سپس در مورد معماری شبکه های آموزش پذیر و توابع هزینه استفاده شده به عنوان دو رکن اصلی شبکه های عمیق مورد استفاده در روش پیشنهادی می پردازیم.

۴ - ۲ - چهارچوب الگوریتم پیشنهادی

در این بخش به شرح روش پیشنهادی پرداخته خواهد شد. شبکه های مولد تخاصمی ایده ای عالی در زمینه فراتفکیک پذیری تصاویر به روش شبکه های عمیق را ارائه می دهند. پژوهش [۳۶] نشان داده است که با فرض استفاده از یک **معماری یکسان** در مدل فراتفکیک پذیر، اعمال *GAN* به شبکه های عمیق منجر به تولید تصاویر خوشایندتری برای انسان در مقایسه با شبکه ی فراتفکیک پذیری عمیقی نظیر *ResNet* می شود که از مدل *GAN* استفاده نکرده است. لذا در رویکرد پیشنهادی از شبکه های مولد تخاصمی به عنوان اساس مدل خود استفاده کردیم. علاوه بر این همانگونه که در بخش ۴-۱ بدان اشاره شد، در فراتفکیک پذیری مبتنی بر ویدیو، علاوه بر فراتفکیک پذیری فریم به فریم، نیاز به دریافت اطلاعاتی اضافی موجود در ترتیب فریم ها داریم که موجب حفظ ساختار پیوستگی زمانی ویدیوی فراتفکیک پذیر شده می شود. در روش پیشنهادی جهت نیل به این هدف، به استخراج اطلاعات مذکور از طریق یک شبکه ی آموزش پذیر شار نوری به نام *FNet* می پردازیم. بلوک دیاگرام روش پیشنهادی در شکل ۴-۱ نشان داده شده است.



شکل ۴-۱-بلوک دیاگرام روش پیشنهادی

روش پیشنهادی متشکل از دو مدل مولد و متمایزگر به عنوان یک شبکه مولد تخاصمی می باشد. در ادامه به شرح کامل الگوریتم پیشنهادی می پردازیم.

۴ - ۲ - ۱ - شبکه مولد

شبکه مولد را می توان صرفاً به عنوان یک شبکه عمیق فراتفکیک پذیر کننده تصویر یعنی *SRNet* مدل سازی نمود که تصویر با کیفیتی را تولید می نماید.

همانگونه که در فصل ۲ بیان شد مسئله فراتفکیک پذیری ویدیو را می توان توسعه ای از مسائل فراتفکیک پذیری تصویر به حساب آورد. به گونه ای که با فراتفکیک پذیری هر فریم و ترکیب مجدد فریم های فراتفکیک شده به خروجی مطلوب که همان ویدیوی فراتفکیک‌شده است رسید. هرچند فراتفکیک پذیری فریم به فریم یک ویدیو موجب تولید اعوجاج های ناخواسته شده و نتایج زمانی پیوسته ای را تولید نمی کند. ضمن آنکه نیاز است روش پیشنهادی بر اساس فراتفکیک پذیری ویدیو (و نه یک تصویر) بهینه سازی گردد قبل از آنکه فریم مورد نظر به شبکه ی فراتفکیک داده شود قصد داریم که اثر فریم های قبلی را به فریمی کنونی اضافه کنیم.

لذا به جای اعمال فراتفکیک پذیری به هر فریم به طور مستقل، می توان از ادغام اطلاعات چندین فریم تصویر با وضوح پایین برای تولید یک فریم با وضوح بالا استفاده نمود و این کار را برای تمامی فریم های ویدیو به ترتیب انجام داد. به عبارت دیگر مسئله فراتفکیک پذیری ویدیو را ما به تعداد زیادی مسئله فراتفکیک پذیری چندفریمی جهت دستیابی به یک فریم با وضوح بالا تبدیل می کنیم.

اگرچه روش فوق اطلاعات به نسبت بالاتری را نسبت به روش فراتفکیک پذیری تک فریمی در اختیار ما قرار می دهد اما ما را با دو چالش جدید مواجه می کند:

۱) هر فریم ورودی چندین بار جهت پردازش فریم های آتی مورد استفاده قرار می گیرد که بار محاسباتی اضافی تحمیل می کند.

۲) هر فریم مستقل از نوع تصویر فریم های قبلی خود تخمین زده می شود و لذا مشکل پیوستگی زمانی در خروجی با وضوح بالا همچنان ممکن است به قوت خود باقی بماند.

در خصوص حل مشکلات فوق، ما در مدل مولد خود از یک رویکرد فریم بازگشتی استفاده کردیم که شامل یک شبکه عمیق آموزش پذیر مبتنی بر شار نوری بوده که علی رغم حفظ پیوستگی زمانی بین فریم های تولید شده با وضوح بالا در هربار محاسبه تنها از یک فریم و نه فریم های ماقبل خود استفاده می کند. مزیت استفاده از این رویکرد بازگشتی را می توان در موارد ذیل خلاصه نمود:

۱) افزایش کیفیت و دقت فراتفکیک پذیری ویدیو با در نظر گرفتن حفظ پیوستگی زمان بین فریم ها
۲) پردازش تنها یک فریم ماقبل در هر دور محاسبه که منجر به کاهش بار محاسباتی الگوریتم می گردد.
۳) در هر دور محاسبه ما از فریم فراتفکیک پذیر شده ماقبل نیز استفاده می کنیم. به علت اینکه فریم فراتفکیک پذیر شده ماقبل حاوی اطلاعات فریم ماقبل خود می باشد، ما به طور غیرمستقیم از اطلاعات فریم های قبل بدون افزایش بار محاسباتی استفاده می کنیم.

لذا نه تنها دقت بهتری نسبت به الگوریتم های هم رده خود دارد بلکه از لحاظ عملکرد از بار محاسباتی کمتری برخوردار می باشد. سوالی که در این بخش ممکن است در ذهن خواننده مطرح شود اینست که دستیابی به اطلاعات شار نوری از طریق روش های کلاسیک میسر است و استفاده از شبکه های عمیق در بلوک شار نوری موجب افزایش پیچیدگی محاسباتی مدل می گردد. توضیح آنکه همانگونه که در بخش ۲-۴-۳ بدان اشاره شد علت استفاده ما از یک شبکه ی عمیق شار نوری کسب ویژگی های مطلوب تر نسبت به روش های محاسبه شار نوری کلاسیک می باشد. لذا هرچند شبکه ی عمیق شار نوری بار محاسباتی اضافه تری را نسبت به روش های کلاسیک ساده محاسبه شار نوری به مدل تحمیل می کند اما موجب کسب نتایج مطلوب تر و مقاوم تری نسبت به تغییرات جهت کسب اطلاعات مرتبط با پیوستگی بین فریم ها خواهد شد. بنابراین شبکه مولد ما شامل دو شبکه آموزش پذیر *FNet* و *SRNet* خواهد بود که شبکه *FNet* اطلاعات حاصل از فریم های قبلی را به فریم کنونی *LR* اضافه نموده و به شبکه *SRNet* جهت تولید تصویر فراتفکیک پذیر شده می دهد.

۴ - ۲ - ۱ - ۱ - الگوریتم شبکه مولد

در این بخش به شرح گام به گام الگوریتم شبکه مولد می پردازیم. خواننده جهت درک بهتر می تواند به شکل ۴-۱ مراجعه نماید.

گام اول-تولید شار نوری *LR*: جهت کسب اطلاعات پیوستگی بین فریم های ویدیو، از شبکه *FNet* جهت دریافت اطلاعات شار نوری استفاده می کنیم. ورودی شبکه آموزش پذیر *FNet*، فریم وضوح پایین کنونی و فریم وضوح پایین ماقبل می باشد. شبکه *FNet* شار نوری بین دو فریم وضوح پایین داده شده را تخمین زده و یک خروجی نرمالیزه شده از نگاشت شار را نتیجه می دهد. روند گفته شده در ۴-۱ بیان شده است:

$$F^{LR} = FNet(I_{t-1}^{LR}, I_t^{LR}) \in [-1,1]^{H\times W\times 2} \qquad -۴-۱$$

به بیان دیگر ۴-۱ مکانی را به هر پیکسل فریم ماقبل براساس موقعیت هرپیکسل در فریم حال حاضر اختصاص می دهد.

گام دوم-تولید شار نوری *HR*: به دلیل اینکه در گام بعد نیازمند ترکیب شار تولیدی با فریم تخمین زده شده قبلی که در ابعاد *HR* می باشد هستیم به افزایش مقیاس شار تولیدی به اندازه فریم تخمین زده شده می پردازیم که دو ورودی یکسان برای ترکیب را در اختیار داشته باشیم. لذا شار تولیدی را با ضریب مقیاس s افزایش ابعاد می دهیم. برای افزایش مقیاس از درون یابی دو خطی به دلیل سادگی محاسباتی استفاده می کنیم. خروجی این بلوک شار تولیدی در مقیاس تصویر وضوح بالا (*HR*) می باشد.

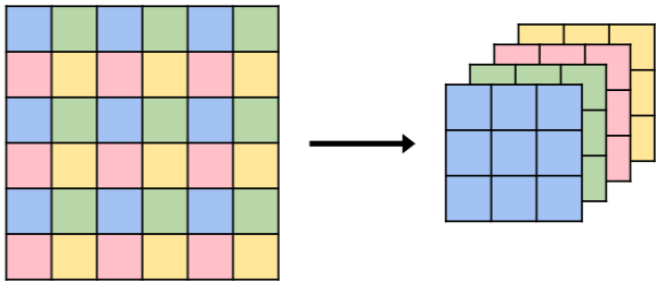
$$F^{HR} = UP(F^{LR}) \in [-1,1]^{sH\times sW\times 2} \qquad -۴-۲$$

گام سوم-پیچش[1] فریم تخمینی قبل و شار نوری *HR*: جهت ادغام اطلاعات بدست آمده از شار نوری با فریم قبلی از عملیاتی به نام پیچش استفاده می کنیم. ادغام این اطلاعات تحت عنوان پیچش بین شار نوری وضوح بالا و فریم تخمین زده شده قبلی و با استفاده از تابع تفاضلی و با استفاده از درون یابی دوخطی انجام می پذیرد. رویکرد استفاده شده در این مرحله بر اساس [۴۰] و مبتنی بر تبدیل معرفی شده در پژوهش [۵۵] انجام شده است.

$$\tilde{I}_{t-1}^{est} = WP(I_{t-1}^{est}, F^{HR}) \qquad -۴-۳$$

گام چهارم-تبدیل فضا به عمق: ورودی بلوک فراتفکیک پذیر *SRNet* تصاویر با ابعاد وضوح پایین می باشند در حالی که خروجی گام سوم فریم تخمینی ماقبل حاوی اطلاعات شار نوری در ابعاد *HR* می باشد. لذا به جای کاهش مقیاس خروجی مرحله قبل یعنی $\tilde{I}_{t-1}^{est}$ که منجر به از دست رفتن اطلاعات می گردد آن را به فضای *LR* با استفاده از تبدیل فضا به عمق نگاشت می دهیم. به عبارت دیگر پیکسل های اضافی در طول و عرض تصویر *HR* را به بعد عمق در یک تصویر *LR* نگاشت می دهیم. خواننده می تواند جهت درک بهتر این عملیات به شکل ۲-۴ مراجعه کند.

$$S_s: [0,1]^{sH \times sW \times C} \rightarrow [0,1]^{H \times W \times s^2 C} \qquad -۴-۴$$



شکل ۲-۴-تبدیل فضا به عمق جهت تبدیل تصویر وضوح بالا به وضوح پایین بدون از دست رفتن اطلاعات

[1] Warp

گام پنجم-جمع فریم تخمینی قبلی و فریم *LR* کنونی: جهت استفاده از اطلاعات حاصل از شار نوری فریم قبلی قبل از فراتفکیک پذیری توسط بلوک *SRNet* ، خروجی های مرحله قبل را به فریم *LR* کنونی یعنی I_t^{LR} تجمیع[1] می کنیم. اعمال این عملیات موجب تولید فریم کنونی *LR* ای بوده که حاوی اطلاعات فریم قبل که با استفاده از شار نوری به دست آمده نیز می باشد.

$$\hat{I}_t^{est} = I_t^{LR} \oplus S_s(\tilde{I}_{t-1}^{est}) \quad -۴-۵$$

گام ششم-فراتفکیک پذیری تصویر: فریم به دست آمده از گام پنجم به شبکه *SRNet* داده می شود. این شبکه به فراتفکیک پذیری فریم کنونی از ویدیو می پردازد که پیش تر اطلاعات حاصل از شار نوری تحت عنوان اطلاعات پیوستگی زمانی ویدیو در آن ادغام شده است.

$$I_t^{est} = SRNet(\hat{I}_t^{est}) \quad -۴-۶$$

۴ - ۲ - ۲ - شبکه متمایزگر

خروجی مرحله قبل یعنی I_t^{est} یک فریم فراتفکیک پذیر شده از ویدیو را نتیجه می دهد. هرچند به دلیل کسب نتایج هرچه بهتر و خوشایندتر برای چشم انسان، شبکه مولد پس از تولید تصویر فراتفکیک پذیر شده آن را به شبکه متمایزگر می دهد. در روش پیشنهادی شبکه متمایزگر D_{θ_D} به نوعی طراحی می شود که با آموزش شبکه مولد بتوان شبکه متمایزگر D_{θ_D} را به نوعی فریب داد که نتواند تصویر فراتفکیک پذیر شده و تصویر واقعی را از یکدیگر تشخیص دهد. به عبارت دیگر وظیفه شبکه متمایزگر که خود یک شبکه آموزش پذیر می باشد اعتبار سنجی تصویر تولید شده شبکه مولد می باشد. نتیجه آنکه خروجی الگوریتم موجب تولید تصاویر بسیار طبیعی و از لحاظ بصری بسیار مطلوب می باشد.

در بخش بعد به شرح معماری شبکه های آموزش پذیر روش پیشنهادی می پردازیم.

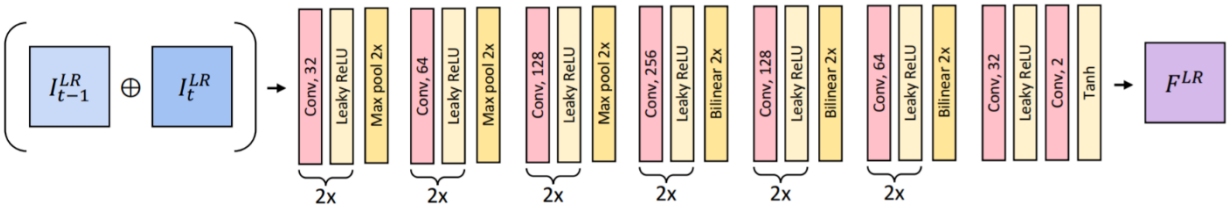
[1] Concatenate

۴ – ۳ – معماری شبکه های آموزش پذیر روش پیشنهادی

همانگونه که در شکل ۳-۷ مشخص است، روش پیشنهادی ما شامل سه بلوک آموزش پذیر به نام های *FNet*، *SRNet* و *Discriminator* می باشد. در ادامه به شرح هر یک از این شبکه ها خواهیم پرداخت:

۴ – ۳ – ۱ – شبکه ***FNet***

این شبکه وظیفه تخمین صحیح شار نوری بین فریم *LR* قبل و بعد را بر عهده دارد که در نتیجه اطلاعات بهتر و دقیق تری را برای شبکه آموزش پذیر بعدی (*SRNet*) فراهم می کند. معماری شبکه طراحی شده به شکل ذیل است. کرنل مورد استفاده 3×3 و *stride=1* در نظر گرفته شده است. ضریب نشتی تابع فعالساز *Leaky ReLu* برابر با ۰.۲ در نظر گرفته شده است. معماری شبکه *FNet* مورد استفاده در این پژوهش بر اساس [۴۰] می باشد.



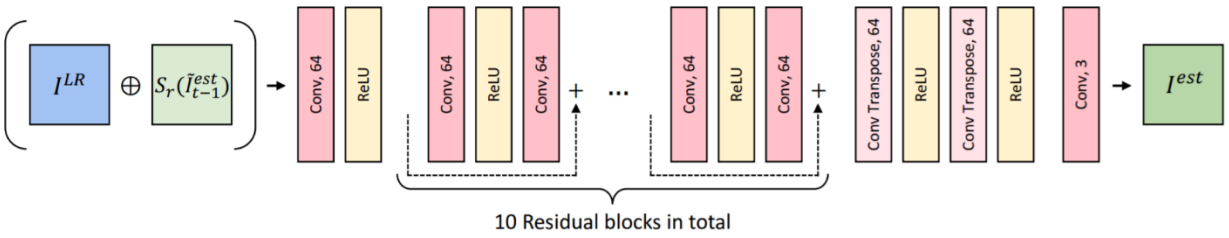
شکل ۳-۴- معماری شبکه *FNet*

۴ – ۳ – ۲ – شبکه ***SRNet***

این شبکه همزمان با شبکه قبل در حال یادگیری بازسازی تصاویر با کیفیت می باشد. در شکل ۳-۷ مشاهده می کنیم که ورودی این شبکه تصویر *LR* فریم کنونی و اطلاعات به دست آمده از شار نوری فریم حال حاضر و فریم قبل است. هرچند شبکه *SRNet* در صورتی که ارتباط شار نوری مناسبی بین فریم *LR* کنونی و $\hat{I}_t^{est}$ پیدا نکند، $\hat{I}_t^{est}$ را نادیده گرفته و تنها به بازسازی فریم حال حاضر اکتفا می کند. به عنوان مثال اولین فریم

از هر ویدیو و یا نواحی که دچار انسداد[1] می شوند نواحی هستند که شبکه *SRNet* تنها به بازسازی فریم حال حاضر می پردازد. روش تشخیص این حالت ها، مقایسه فرکانس های پایین $\hat{I}_{t-1}^{est}$ با فریم کنونی *LR* یعنی I_t^{LR} می باشد. در صورتی که تطبیقی بین این دو پیدا نشود، شبکه، $\hat{I}_{t-1}^{est}$ را نادیده گرفته و صرفاً فریم کنونی را فراتفکیک پذیری می کند. زمانی که کار آموزش شبکه به پایان رسید می تواند روی هر ویدیو و با هر اندازه ای اعمال گردد.

شبکه *SRNet* برای افزایش مقیاس ۴ برابری از یک طراحی شبکه همگشتی تمام متصل استفاده می کند که شامل ۱۰ بلوک باقیمانده میانی است. کرنل مورد استفاده 3×3 و $stride=1$ می باشد. ایده طراحی معماری شبکه *SRNet* بر مبنای روش پیشنهاد شده در [۴۰] می باشد.



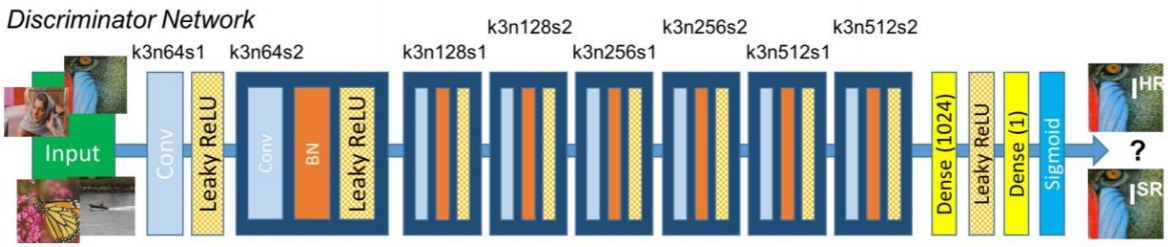
شکل ۴-۴- معماری شبکه *SRNet*

۴ - ۳ - ۳ - معماری شبکه متمایزگر

جهت توسعه شبکه متمایزگر ما از معماری مقاله [۳۶] کمک گرفته ایم. به گونه ای که ضریب نشتی $ReLU=0.2$ در نظر گرفته شده و از *max-pooling* در حین آموزش اجتناب نموده ایم. این شبکه شامل ۸ لایه همگشتی بوده و از کرنل 3×3 و با ضریب افزایشی ۲ (از ۶۴ به ۵۱۲) در شبکه *VGG* استفاده می کند. ضمن اینکه در هر مرحله از واگشت گسترده (با گام بیش از ۱) استفاده می شود که وضوح تصاویر تولید شده در هر گام که

[1] Occluded Areas

ویژگی ها دو برابر می شود را کاهش می دهد. در نهایت ۵۱۲ ویژگی نگاشت داده شده به دولایه متراکم[1] داده شده و در نهایت از یک تابع فعال ساز *sigmoid* برای به دست آوردن یک مقدار احتمال جهت طبقه بندی استفاده می کند.



شکل ۵-۴- معماری شبکه *Discriminator*

۴ - ۴ - توابع هزینه

در الگوریتم پیشنهادی جهت بازسازی فریم های *HR* نیازمند تابع و یا توابع هزینه بهینه هستیم. تابع هزینه مرسومی که در فراتفکیک پذیری مورد استفاده قرار می گیرد مجموع مربعات خطا می باشد. هرچند به دلیل ناتوانی تابع هزینه مذکور در بازیابی بافت های با فرکانس بالا از توابع هزینه دیگری نیز استفاده می کنیم. تابع هزینه کل برابر مجموع وزن دار از توابع هزینه‌ای ست که در ادامه شرح خواهیم داد.

۴ - ۴ - ۱ - تابع هزینه مجموع مربعات خطا

کمینه نمودن میانگین مربعات خطا (*MSE*) بین تصویر تخمینی و تصویر مرجع (*Ground Truth*) متداول ترین روش مورد استفاده در فراتفکیک پذیری است. این تابع هزینه می تواند به شکل ذیل به شکل ریاضی نوشت:

[1] Dense layer

$$L_{MSE} = \frac{1}{s^2WH}\sum_{x=1}^{sW}\sum_{y=1}^{sH}(I_{x,y}^{HR} - I_{x,y}^{SR})^2 \quad -۷-۴$$

توابع هزینه حساس به پیکسل نظیر *MSE* علی رغم افزایش *PSNR*، در بازیابی جزئیات فرکانس بالای تصویر ناتوان هستند. لذا حداقل نمودن میزان *MSE* بافت تصاویر را نرم کردن و لذا موجب کاهش دید ادراکی تصویر می گردد. نتیجه آنکه *MSE* به تنهایی نمی تواند تابع هزینه ما باشد و باید توابع دیگری که بتواند بر مشکل مذکور غلبه کند تعریف گردد.

۴ - ۴ - ۲ - تابع هزینه ادراکی[1]

جهت غلبه بر مشکل مذکور از تابع هزینه دیگری به نام تابع هزینه ادراکی استفاده می کنیم که بر اساس نگاشت ویژگی های شبکه *VGG* به دست می آید. [۳۶] نشان داده که تابع هزینه ادراکی از ثبات بهتری نسبت به فضای پیکسلی در مقایسه با *MSE* برخوردار بوده و لذا از محو شدگی تصویر جلوگیری می کند. این تابع هزینه مبتنی بر لایه های فعال ساز *ReLU* شبکه *VGG19* می باشد. بر این اساس $\emptyset_{i,j}$ نمایانگر نگاشت ویژگی به دست آمده از j امین پیچش قبل از i امین لایه *Maxpooling* در شبکه *VGG19* می باشد. در نهایت تابع هزینه را بر اساس فاصله اقلیدسی بین ویژگی های تصویر بازسازی شده I_t^{SR} و تصویر مرجع I_t^{HR} تعریف می کنیم. ۸-۴ این مفهوم را به شکل ریاضی نشان می دهد:

$$L_{Perceptual} = \frac{1}{W_{i,j}H_{i,j}}\sum_{x=1}^{W_{i,j}}\sum_{y=1}^{H_{i,j}}(\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(I^{SR})_{x,y})^2 \quad -۸-۴$$

[1] Perceptual loss function

۴ - ۴ - ۳ - تابع هزینه تخاصمی[1]

با افزودن بخش مولد[2] شبکه *GAN* به تابع هزینه می توان به طبیعی کردن تصویر کمک نمود. این کار با فریب شبکه متمایز کننده[3] رخ می دهد. تابع هزینه مولد[4] بر مبنای مجموع احتمال شبکه ی تمایز بر روی تمامی نمونه ها بیان می گردد. این مفهوم را به شکل ریاضی در ۹-۴ بیان می کنیم.

$$L_{Adverarial} = \sum_{n=1}^{N} -logD_{\phi_D}(G_{\phi_G}(I^{LR})) \qquad -۹-۴$$

در این بخش $D_{\phi_D}(G_{\phi_G}(I^{LR}))$ احتمال اینکه شبکه متمایزگر تصویر بازسازی شده را به عنوان یک تصویر وضوح بالا تشخیص دهد را بیان می کند. خواننده توجه داشته باشد که جهت ساده سازی ، معادله ۱۰-۴ به شکل ۹-۴ نوشته شده است:

$$\log\left[1 - logD_{\phi_D}\left(G_{\phi_G}(I^{LR})\right)\right] \qquad -۱۰-۴$$

۴ - ۴ - ۴ - تابع هزینه *TV*

این تابع از مجموع مطلق تفاضل مقادیر پیکسل همسایه در دو مسیر افقی و عمودی تصویر به دست می آید. به دلیل اینکه تابع هزینه *TV* نویز موجود در تصویر ورودی را نشان می دهد، حداقل نمودن آن به عنوان بخشی از تابع هزینه کل می تواند موجب کاهش نویز موجود در تصویر و نرم تر کردن آن گردد. تابع هزینه *TV* را می توان به شکل ذیل به صورت ریاضی نوشت:

$$L_{TV} = \frac{1}{WH}\sum_{i=0}^{W}\sum_{j=0}^{H}\sqrt{(G_{\theta_G}LR_{t_{i,j+1,k}} - G_{\theta_G}LR_{t_{i,j,k}})^2 + (G_{\theta_G}LR_{t_{i+1,j,k}} - G_{\theta_G}LR_{t_{i,j,k}})^2} \qquad -۱۱-۴$$

[1] Adversarial loss function
[2] Generative
[3] Discriminator
[4] Generative loss function

۴ - ۴ - ۵ - تابع هزینه *Flow*

به دلیل اینکه شار نوری موجود در تصویر اصلی با وضوح بالا (مرجع) موجود نیست. تابع هزینه *Flow* را به صورت میانگین مربع خطای بین تصویر با وضوح پایین I_t^{LR} و پیچش بین شار وضوح پایین F^{LR} و تصویر با وضوح پایین فریم قبل I_{t-1}^{LR} به دست می آوریم.

$$L_{flow} = ||WP(I_{t-1}^{LR}, F^{LR}) - I_t^{LR}||_2^2 \qquad -۱۲-۴$$

۴ - ۴ - ۶ - تابع هزینه کل

همانگونه که بیان شد تابع هزینه کل برابر مجموع وزن دار از توابع هزینه فوق می باشد. لذا خواهیم داشت:

$$L_{total} = \alpha \times L_{MSE} + \beta \times L_{Perceptional} + \gamma \times L_{Advarsial} + \delta \times L_{TV} + \epsilon \times L_{flow} \qquad -۱۳-۴$$

در فصل آینده به پیاده سازی و ارزیابی نتایج خواهیم پرداخت.

فصل ۵
پیاده سازی و ارزیابی نتایج

۵ – ۱ – مقدمه

در این فصل به پیاده سازی روش پیشنهادی ارائه شده و بررسی نتایج آن خواهیم پرداخت. همانطور که در فصل ۳ اشاره شد، روش پیشنهادی ما مبتنی بر شبکه های عمیق بوده و شامل دو مرحله یادگیری و آزمایش می باشد. در این فصل به معرفی سخت افزارها و نرم افزارهای استفاده شده برای پیاده سازی روش پیشنهادی خواهیم پرداخت. در ادامه به معرفی پارامترهای استفاده شده، نتایج به دست آمده و مقایسه روش پیشنهادی با سایر روش های هم رده خود خواهیم پرداخت.

۵ – ۲ – سخت افزار و نرم افزار استفاده شده

به دلیل اینکه روش پیشنهادی ما شامل ۳ شبکه آموزش پذیر می باشد لذا آموزش الگوریتم پیشنهادی نیازمند سخت افزار قدرتمند و پیشرفته می باشد.

۵ - ۲ - ۱ - سخت افزار

Graphic: RTX 2060 Super

Ram: 32GB

CPU: Core i5 9600K

۵ - ۲ - ۲ - نرم افزار

پیاده سازی الگوریتم از طریق زبان پایتون و کتابخانه یادگیری ماشین *Pytorch* در نرم افزار *PyCharm* و سیستم عامل لینوکس اوبونتو[1] انجام گرفته است.

Language: Python 3.7

IDE: PyCharm 2020.2

Machine Learning Library: Pytorch

OS: Linux Ubuntu

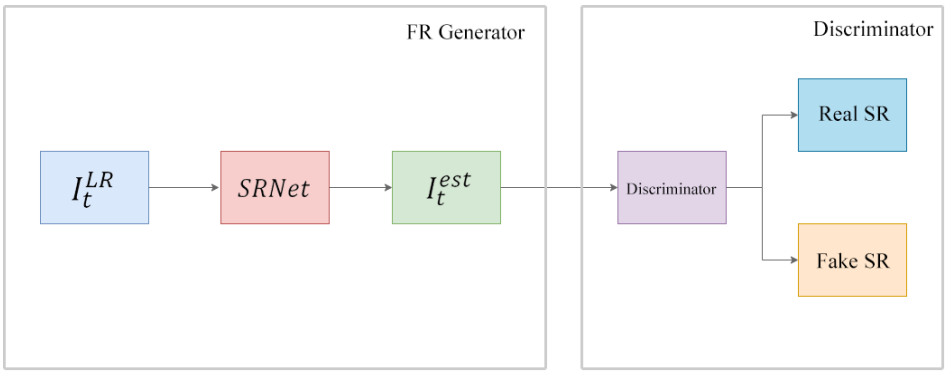
۵ – ۳ – مدل های شبیه سازی شده

جهت بررسی بهتر در این پژوهش، مدل پیشنهادی با دو مدل دیگر که در مقالات [۳۶] و [۴۰] آورده شده اند مقایسه شده و نتایج آن در ادامه آورده خواهد شد.

۵ - ۳ - ۱ - فراتفکیک پذیری تصویر توسط شبکه *GAN (SRGAN)*

در این مدل، در شبکه مولد (*G*)، شبکه آموزش پذیر *FNet* حذف شده و فریم کنونی مستقیماً توسط شبکه *SRNet* تخمین زده می شود. در نهایت، فریم تخمینی به شبکه متمایزگر (*D*) داده می شود. این مدل بر مبنای مقاله [۳۶] شبیه سازی و توسعه داده شده است. شکل ۱-۵ مدل مربوطه را به شکل بهتری توصیف می کند. مدل مذکور در ادامه پژوهش *SRGAN* نامیده می شود.

[1] Linux Ubuntu



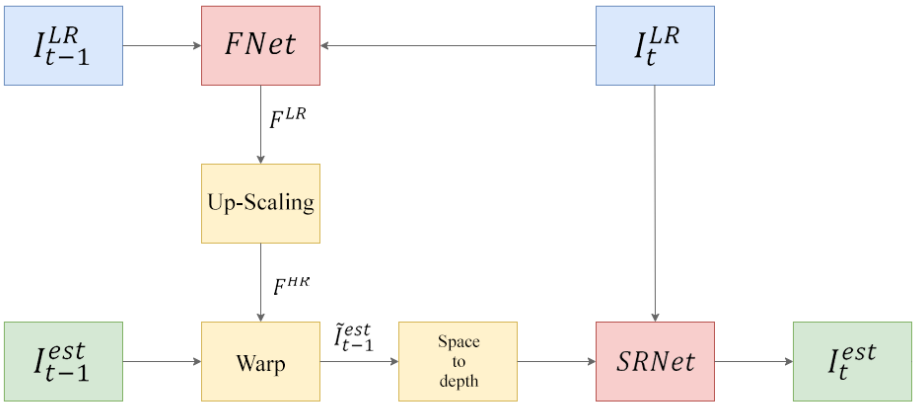
شکل ۱۵-۵- نمایش بلوکی مدل *SRGAN*

همانطور که در شکل ۱-۵ هم مشاهده می شود، مدل مربوطه مبتنی بر فراتفکیک پذیری تک فریم (فراتفکیک پذیری تصویر) بوده و در صورت استفاده در فراتفکیک پذیری ویدیو، هر فریم فراتفکیک پذیر شده حاوی اطلاعات فریم های ماقبل خود نخواهد بود.

معماری شبکه های آموزش پذیر *SRNet* و *Discriminator* در مدل *SRGAN* جهت مقایسه عادلانه، همانند مدل پیشنهادی این پایان نامه در نظر گرفته و بدون تغییر مانده است.

۵-۳-۲- فراتفکیک پذیری ویدیو به کمک شبکه ی *FNet (FRVSR)*

در این مدل این بار شبکه مولد تخاصمی (*GAN*) را حذف نموده و صرفاً از شبکه عمیق *FNet* جهت فراتفکیک پذیری ویدیو استفاده می کنیم. شبیه سازی مدل مذکور بر مبنای مقاله [۴۰] انجام شده است. این شبکه در ادامه *FRVSR* نامیده می شود. شکل ۲-۵ ، مدل مربوطه را به شکل بلوکی نشان می دهد.

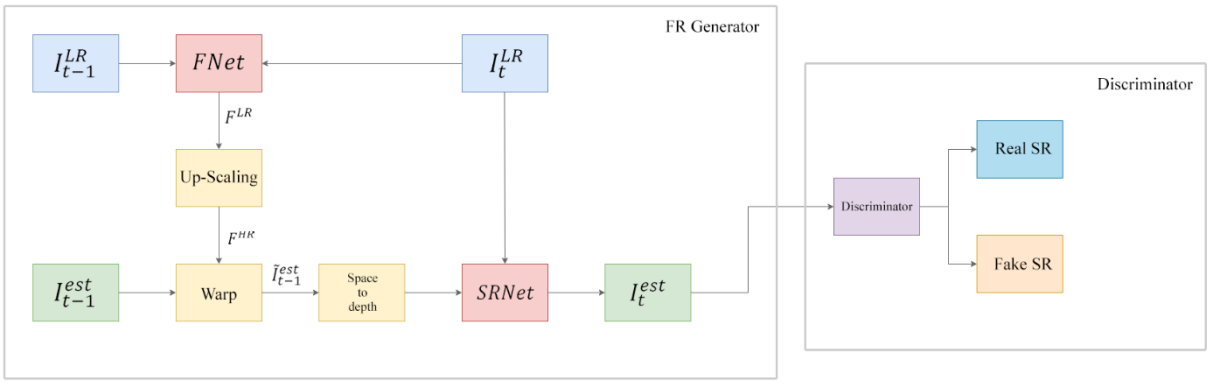


شکل ۲-۵- نمایش بلوکی مدل *FRVSR*

برخلاف مدل *SRGAN*، مدل *FRVSR* از فریم ماقبل خود نیز استفاده نموده و هر فریم حاوی اطلاعات فریم ماقبل خود می باشد (فراتفکیک پذیری ویدیو). معماری شبکه *FNet* همانند مدل پیشنهادی پایان نامه در نظر گرفته شده و بدون تغییر مانده است.

۵ - ۳ - ۳ - فراتفکیک پذیری ویدیو با استفاده از شبکه *GAN* و *FNet* (*FRGANVSR*)

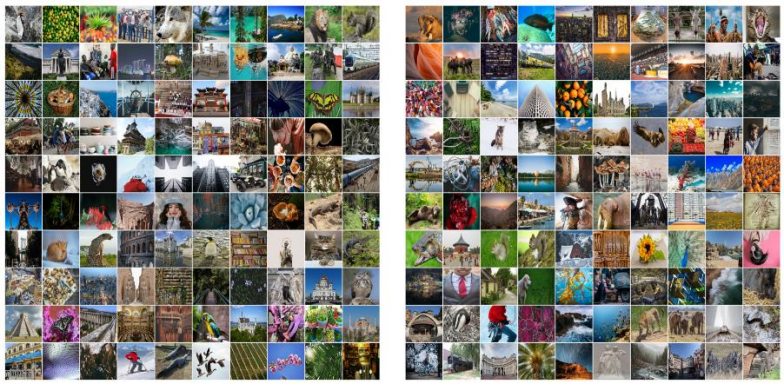
مدل پیشنهادی این پژوهش بوده که از ترکیب شبکه های *FNet* و *GAN* جهت فراتفکیک پذیری استفاده می کند. مدل پیشنهادی *FRGANVSR* نامیده می شود و جهت فراتفکیک پذیری هر فریم از اطلاعات ما قبل خود نیز استفاده می کند (فراتفکیک پذیری ویدیو). مدل *FRGANVSR* در شکل ۳-۵ جهت مقایسه دوباره آورده شده است.



شکل ۳-۵- نمایش بلوکی مدل *FRGANVSR*

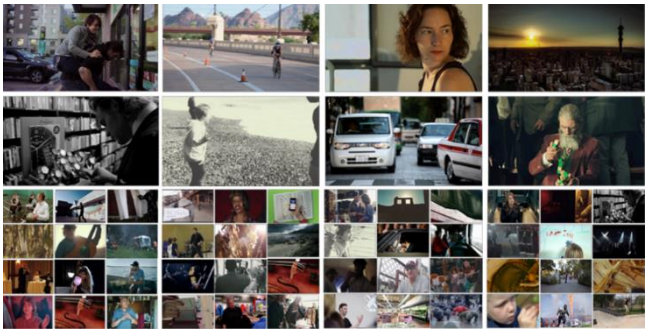
۵ - ۴ - مجموعه داده

همانطور که پیش تر گفته شد مدل *SRGAN* یک مدل فراتفکیک پذیری تک فریمی بوده و لذا نیاز است که از مجموعه داده تصویر جهت آموزش مدل استفاده می کنیم. در شبیه سازی انجام شده برای آموزش مدل *SRGAN* از مجموعه داده *DIV2K Dataset* [۵۶] استفاده نمودیم که شامل ۸۰۰ تصویر با وضوح بالا می باشد. فریم های متناظر *LR* در حین یادگیری تولید می گردند.



شکل ۴-۵- مجموعه داده *DIV2K*

از طرف دیگر به دلیل آنکه مدل های *FRVSR* و *FRGANVSR* مدل های مبتنی بر فراتفکیک پذیری ویدیو هستند جهت آموزش و ارزیابی این دو مدل از مجموعه داده *Vimeo_TOFlow* استفاده نموده که در پروژه *TOFlow* آزمایشگاه علوم کامپیوتر و هوش مصنوعی دانشگاه *MIT* انجام شده است [۵۷]. که حاوی ۷۸۰۰ ویدیو فریم *HR* نمونه برداری شده از سایت *Vimeo.com* می باشد. هر ویدیو شامل ۷ فریم ترتیبی می باشد. در حین روند یادگیری فریم های متناظر *LR* با مقیاس $4\times$ تولید می گردد.



شکل ۱-۵- مجموعه داده‌ی *Vimeo_toflow*

۵ – ۵ – فراپارامترها[1]

در پروسه یادگیری الگوریتم از نرخ یادگیری 1×10^{-5} و اندازه دسته[2] ۲ استفاده شده است. مدل SRGAN در ۱۰۰ دوره[3] ، مدل *FRVSR* در ۲۰ دوره و مدل *FRGANVSR* در ۱۰ دوره آموزش دیده اند. جدول ذیل پارامترهای مورد استفاده در شبیه سازی ها را خلاصه نموده است. انتخاب نرخ یادگیری 1×10^{-5} مبتنی بر پژوهش های مرجع [۳۶, ۴۰] بوده است. ضمن آنکه اندازه دسته بیش از ۲ باعث بروز خطای حافظه گرافیک می شد.

[1] Hyper Parameter
[2] Batch size
[3] Epoch

جدول ۱-۵-پارامترهای استفاده شده در مدل های شبیه سازی شده

نوع پارامتر/مدل	*FRGANVSR*	***FRVSR***	***SRGAN***
نرخ یادگیری (*Learning rate*)	1×10^{-5}	1×10^{-5}	1×10^{-5}
اندازه دسته (*Batch size*)	۲	۲	۲
تعداد دوره (*Epochs*)	۱۰	۲۰	۱۰۰

۵ - ۵ - ۱ - انتخاب ضرایب توابع هزینه

همانطور که فصل ۳ بیان گردید تابع هزینه کل مجموعه به صورت ذیل بیان می گردد:

$$L_{total} = \alpha \times L_{MSE} + \beta \times L_{Perceptional} + \gamma \times L_{Advarsial} + \delta \times L_{TV} + \epsilon \times L_{flow} \quad -۱-۴$$

ضرایب توابع هزینه استفاده شده در پیاده سازی روش پیشنهادی در جدول ۲-۴ آورده شده است. انتخاب ضرایب به دست آمده بر اساس ضرایب بهینه انتخابی که به صورت تجربی در پژوهش های [۳۶, ۴۰] انتخاب شده است.

جدول ۲-۵- ضرایب مورد استفاده در توابع هزینه روش پیشنهادی

نوع تابع هزینه	نماد	مقدار
ضریب تابع هزینه *MSE*	α	1
ضریب تابع هزینه ادراکی	β	6×10^{-3}
ضریب تابع هزینه متخاصم	γ	1×10^{-3}
ضریب تابع هزینه *TV*	δ	2×10^{-8}
ضریب تابع هزینه شار نوری	ϵ	1×10^{-4}

۵ - ۶ - نتایج فاز یادگیری

یادگیری بر اساس فراپارامترهای گفته شده در بخش ۴-۵ و برای ۳ مدل مذکور انجام شده است.

داده های به دست آمده به ازای ۱۰ دوره برای سه مدل به شکل ذیل می باشد:

جدول ۵-۳-نتایج فاز یادگیری برای مدل *SRGAN*

Epoch	G_{loss}	D_{loss}
1	*0.04057*	*0.8480*
10	*0.010908*	*1.003087*
20	*0.00788*	*0.99724*
30	*0.006952*	*1.001077*
40	*0.007145*	*0.998149*
50	*0.006208*	*0.999335*
60	*0.005887*	*0.996082*
70	*0.006482*	*0.996494*
80	*0.00589*	*1.000897*
90	*0.005846*	*1.003689*
100	*0.00587*	*0.995858*

جدول ۵-۴-نتایج فاز یادگیری برای مدل *FRVSR*

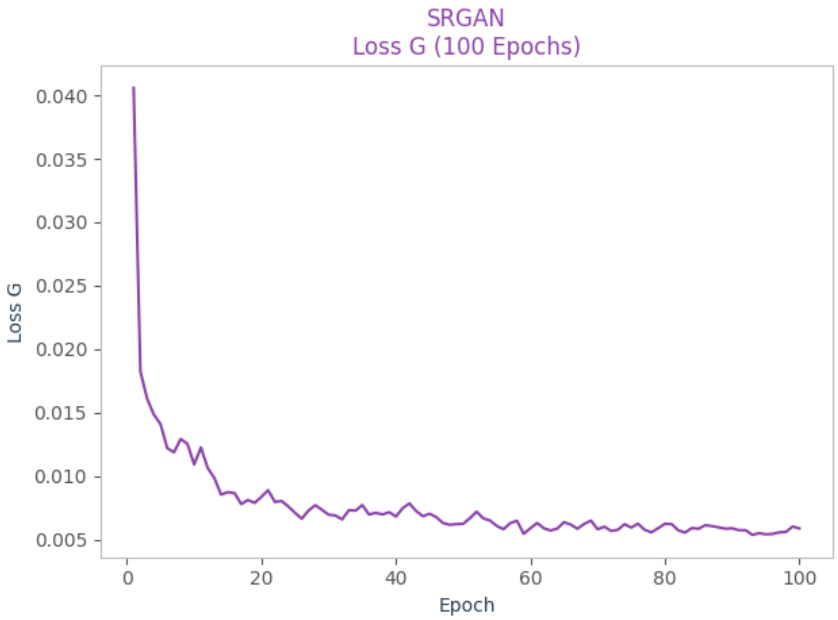
Epoch	***Train Loss***	***Validation Loss***
1	*0.141232*	*0.0822*
3	*0.047753*	*0.0426*
4	*0.039973*	*0.0366*
6	*0.03267*	*0.0311*
8	*0.028749*	*0.0278*
10	*0.026654*	*0.0261*
12	*0.02428*	*0.0238*
14	*0.02286*	*0.0225*
16	*0.021911*	*0.0235*
18	*0.021188*	*0.0209*
20	*0.020703*	*0.0206*

جدول ۵-۵-نتایج فاز یادگیری برای مدل *FRGANVSR*

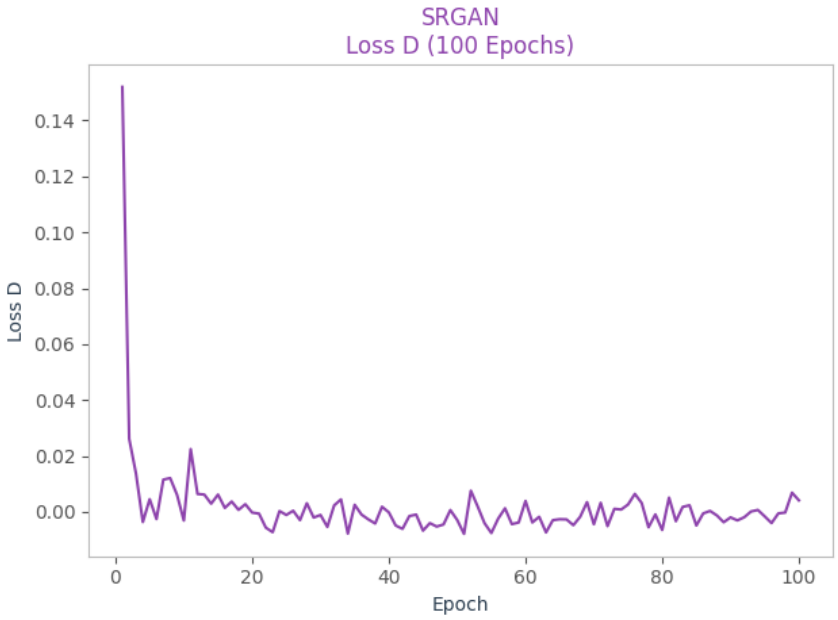
$Epoch$	G_{loss}	D_{loss}
1	*0.012913*	*0.052605*
2	*0.004902*	*0.013777*
3	*0.003921*	*0.018282*
4	*0.00329*	*0.010963*
5	*0.00299*	*0.010437*
6	*0.002827*	*0.007475*
7	*0.002718*	*0.004358*
8	*0.002626*	*0.004852*
9	*0.002582*	*0.002541*
10	*0.002515*	*0.002892*

۵ - ۶ - ۱ - نتایج یادگیری مدل *SRGAN*

مدل های مبتنی بر *GAN* شامل دو بلوک جداگانه در آموزش هستند و لذا مقادیر آنها به شکل مجزا محاسبه شده است. مقادیر هزینه بلوک متمایزگر D_{loss} و هزینه بلوک مولد G_{loss} برای مدل *SRGAN* در ۱۰۰ دوره انجام شده و نتایج آن به شکل ذیل ارائه می گردد.



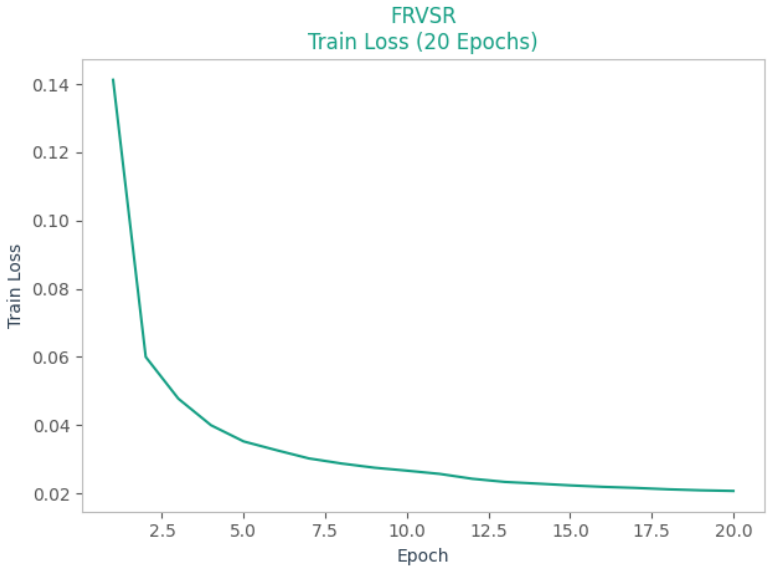
شکل ۵-۵-مقدار G_{loss} به دست آمده به ازای ۱۰۰ دوره در مدل *SRGAN*



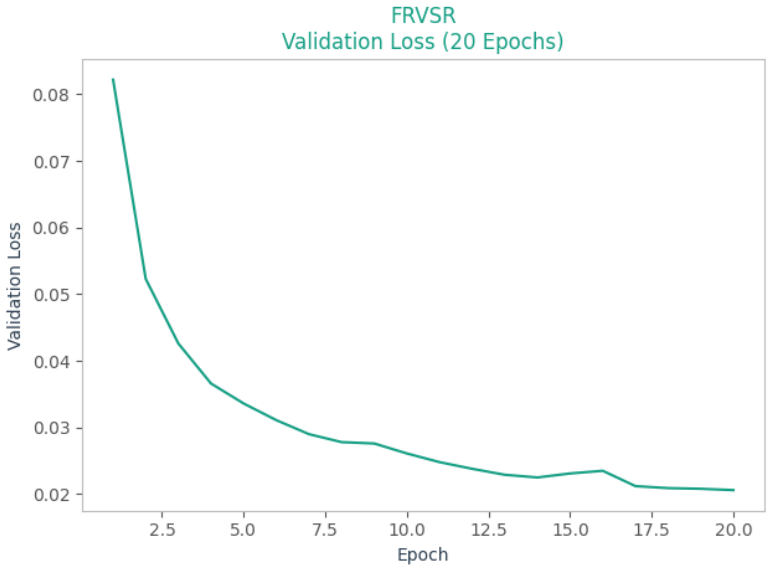
شکل ۶-۵- مقدار D_{loss} به دست آمده به ازای ۱۰۰ دوره در مدل *SRGAN*

۵ - ۶ - ۲ - نتایج یادگیری مدل *FRVSR*

در این شبیه سازی، تابع هزینه کل با عنوان *Train Loss* و *Val Loss* برای مدل *FRVSR* و در ۲۰ دوره یادگیری در شکل های ۵-۷ و ۵-۸ رسم شده است.



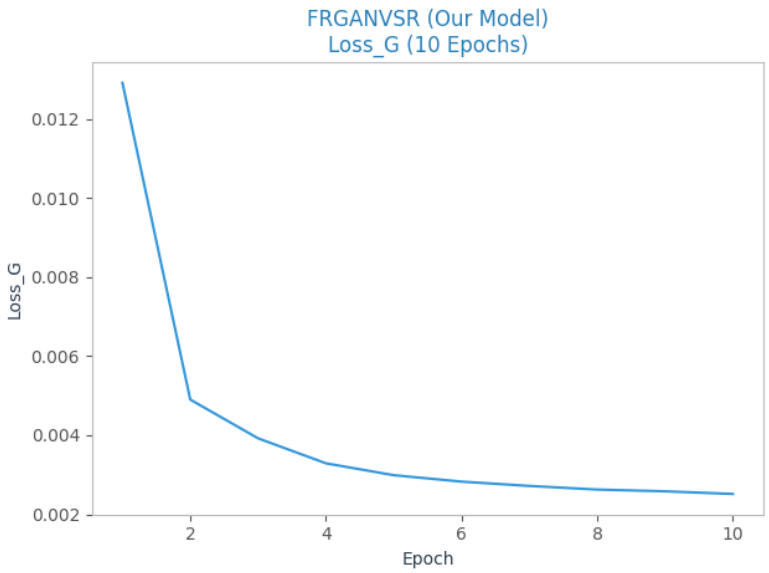
شکل ۵-۷- مقدار *Train Loss* به دست آمده در مدل *FRVSR*



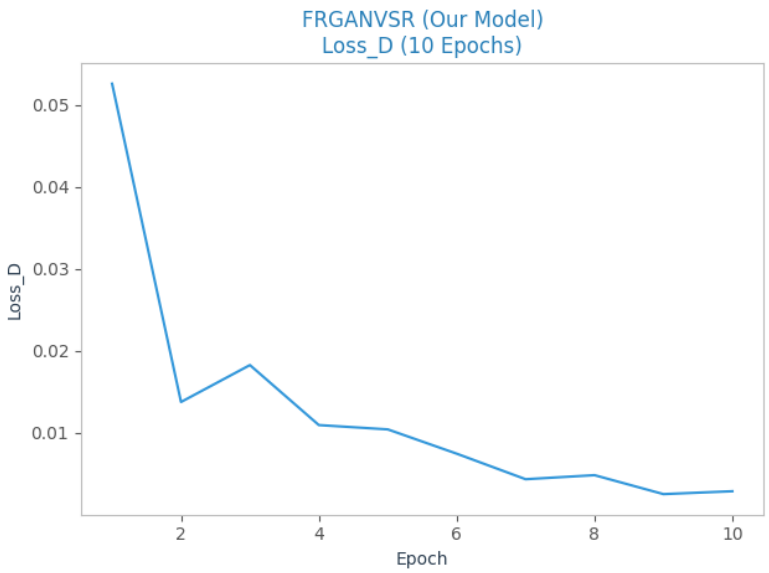
شکل ۵-۸-مقدار *Validation Loss* به دست آمده در مدل *FRVSR*

۵ - ۶ - ۳ - نتایج یادگیری مدل *FRGANVSR*

مدل پیشنهادی (*FRGANVSR*) مبتنی بر *GAN* بوده و مقادیر هزینه بلوک متمایزگر D_{loss}، هزینه بلوک مولد G_{loss}، امتیاز بلوک متمایزگر D_{score}، امتیاز بلوک مولد G_{score} در ۱۰ دوره انجام شده است.



شکل ۹-۵-مقدار G_{loss} به دست آمده به ازای ۱۰ دوره در مدل *FRGANVSR*



شکل ۱۰-۵- مقدار D_{loss} به دست آمده به ازای ۱۰ دوره در مدل *FRGANVSR*

۵ – ۷ – نتایج فاز ارزیابی

معیار *PSNR* و *SSIM* دو معیاری هستند که به صورت گسترده در سنجش کیفیت تصویر مورد استفاده قرار می گیرند. از این رو نتایج فاز ارزیابی بر اساس این دو معیار مورد سنجش قرار گرفته است. معیار *PSNR* به تنهایی نمی تواند معیاری برای کیفیت مدل باشد. از طرف دیگر معیار *SSIM* که بر اساس سه فاکتور روشنایی، تباین و ساختار تصویر کیفیت تصویر را مورد سنجش قرار می دهد می تواند سنجش بهتری جهت بررسی کیفیت بصری تصویر باشد.

جدول ۵-۶-نتایج فاز ارزیابی برای مدل *SRGAN*

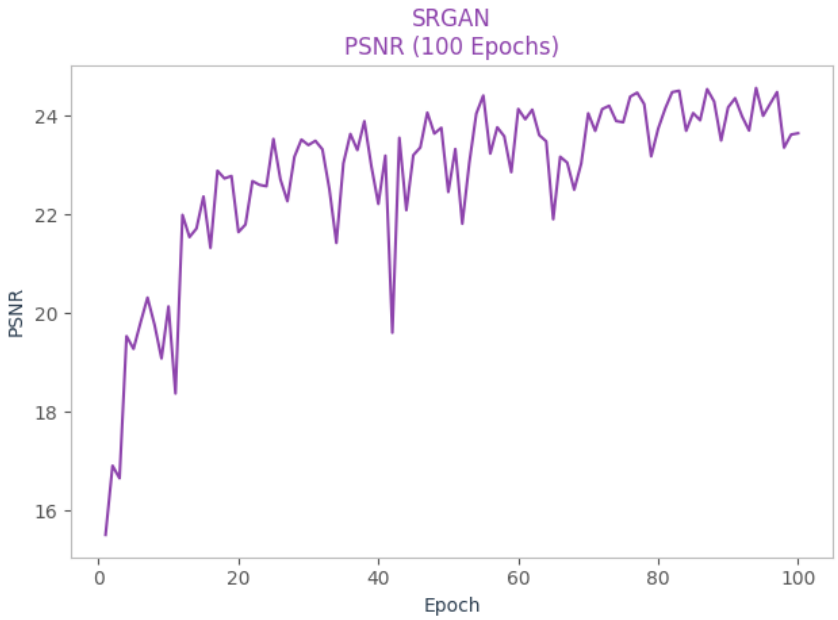
Epoch	***PSNR***	***SSIM***
1	*15.5088*	*0.409119*
10	*20.13852*	*0.594258*
20	*22.78292*	*0.664766*
30	*0.513835*	*0.695989*
40	*22.98261*	*0.725179*
50	*23.7602*	*0.733602*
60	*22.86088*	*0.731118*
70	*23.03681*	*0.732311*
80	*23.74139*	*0.731898*
90	*24.1747*	*0.734163*
100	*23.64915*	*0.736538*

جدول ۵-۷-نتایج فاز ارزیابی برای مدل *FRVSR*

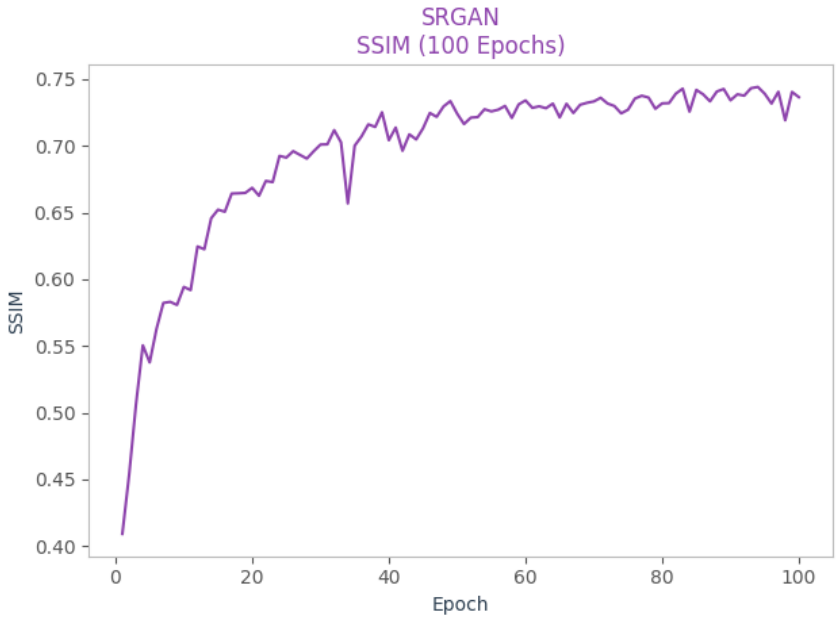
Epoch	*PSNR*	*SSIM*
1	*21.2933*	*0.6829*
2	*24.6569*	*0.7472*
4	*26.6077*	*0.7954*
6	*27.4255*	*0.8229*
8	*27.9186*	*0.8393*
10	*28.1787*	*0.8474*
12	*28.3379*	*0.8507*
14	*28.472*	*0.855*
16	*28.608*	*0.857*
18	*28.6853*	*0.8624*
20	*28.6935*	*0.8638*

جدول ۵-۸-نتایج فاز ارزیابی برای مدل *FRGANVSR*

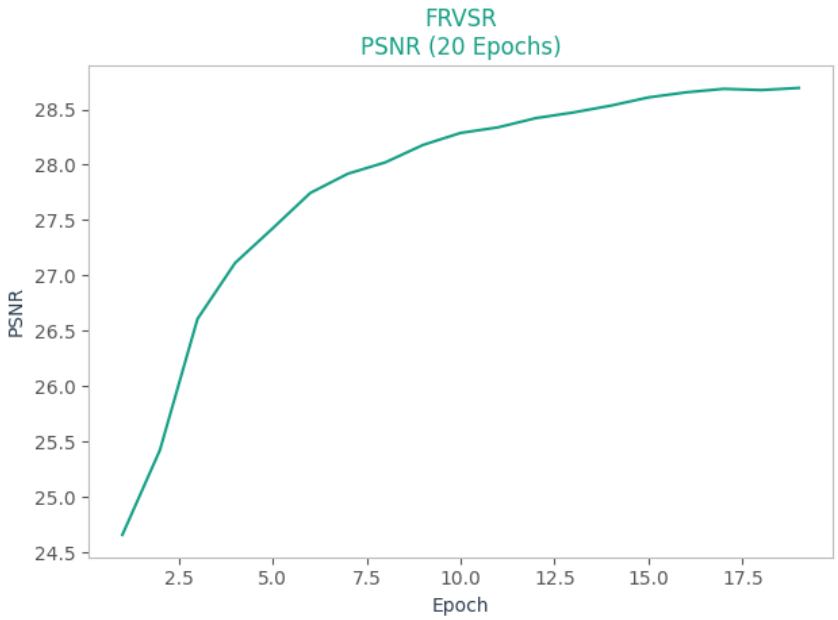
Epoch	*PSNR*	*SSIM*
1	*24.11415*	*0.7272*
2	*25.45516*	*0.7809*
3	*26.96382*	*0.8091*
4	*27.33324*	*0.8244*
5	*27.73934*	*0.8335*
6	*27.93827*	*0.8362*
7	*28.12177*	*0.8466*
8	*28.20049*	*0.8497*
9	*28.35041*	*0.8510*
10	*28.42268*	*0.8541*



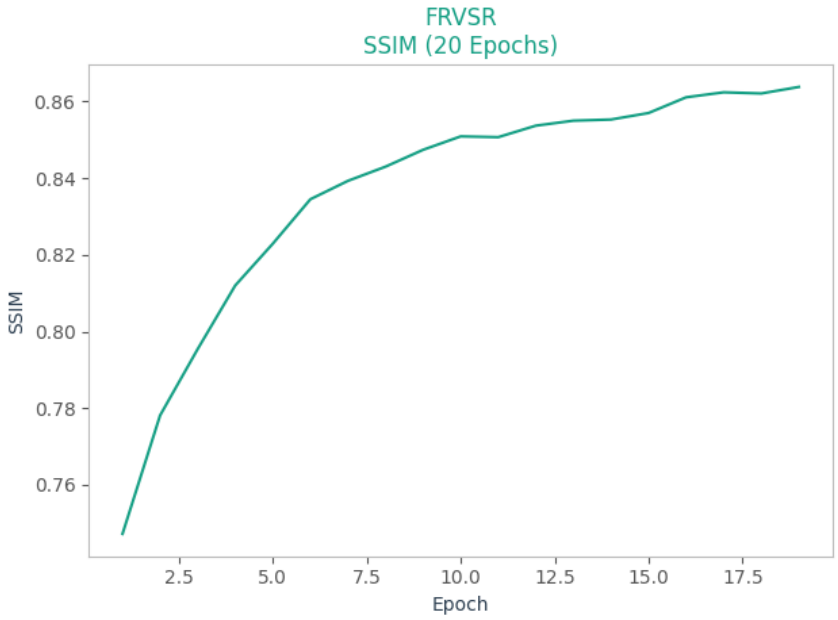
شکل ۱۱-۵-نتایج *PSNR* در فاز ارزیابی مدل *SRGAN*



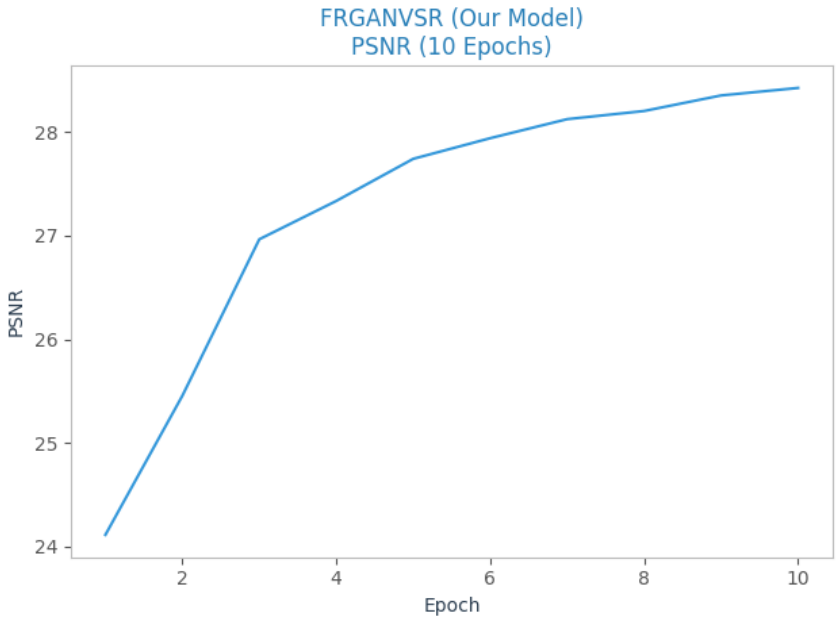
شکل ۱۲-۵- نتایج *SSIM* در فاز ارزیابی مدل *SRGAN*



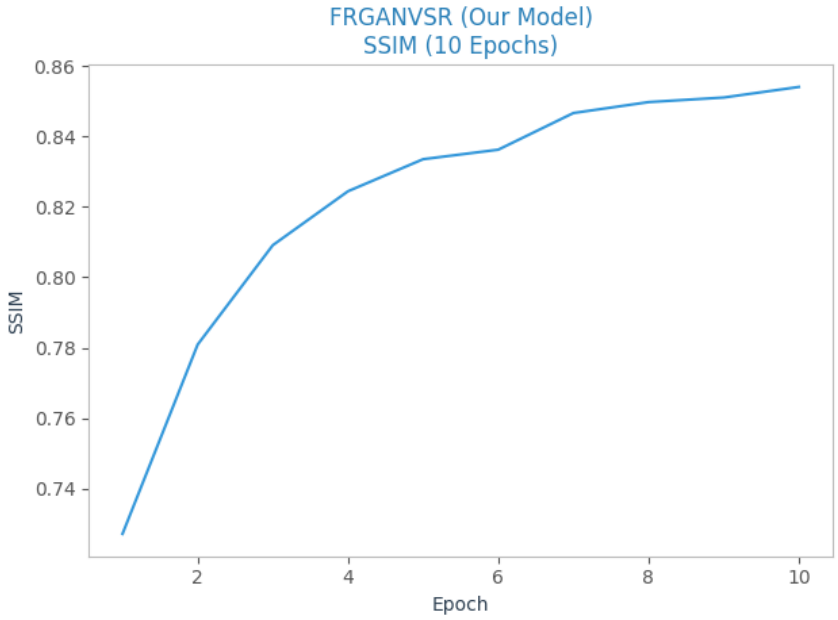
شکل ۱۳-۵- نتایج *PSNR* در فاز ارزیابی مدل *FRVSR*



شکل ۱۴-۵-نتایج *SSIM* در فاز ارزیابی مدل *FRVSR*



شکل ۱۵-۵-نتایج *PSNR* در فاز ارزیابی مدل *FRGANVSR*



شکل ۱۶-۵-نتایج *SSIM* در فاز ارزیابی مدل *FRGANVSR*

۵ - ۷ - ۱ - نتیجه گیری ارزیابی

نتایج *PSNR* و *SSIM* نهایی در ادامه جهت بررسی آورده شده است. جهت مقایسه عادلانه به دلیل آنکه مدل *FRGANVSR* طی ۱۰ دوره آموزش داده شده است و از طرف دیگر مدل *FRVSR* توسط مجموعه داده یکسان طی ۲۰ دوره آموزش داده شده است. نتایج برای مدل *FRVSR* را به دو قسمت *FRVSR10* به معنای نتایج برای ۱۰ دوره (دوره یکسان با مدل *FRGANVSR*) و *FRVSR20* به معنای نتایج برای ۲۰ دوره تقسیم می کنیم.

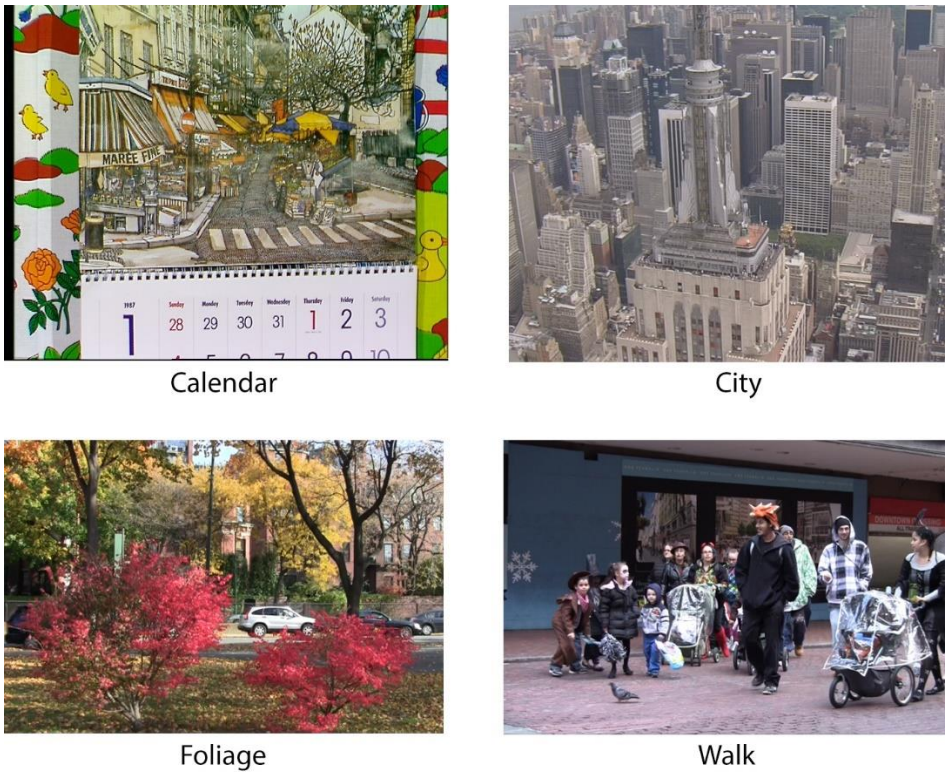
جدول ۵-۹-مقایسه نتایج فاز ارزیابی مدل ها

Model	*PSNR*	*SSIM*	*Epochs*
SRGAN	*23.6491*	*0.7365*	*100*
FRVSR10	*28.1787*	*0.8474*	*10*
FRVSR20	*28.6935*	*0.8638*	*20*
FRGANVSR	*28.4227*	*0.8541*	*10*

همانطور که در جدول ۵-۹ مشاهده می کنیم، مدل پیشنهادی (*FRGANVSR*) در تعداد دوره یکسان با *FRVSR* چه از لحاظ *PSNR* و چه از لحاظ *SSIM* نسبت به سایر مدل ها برتری دارد. هرچند در تعداد دوره۲۰، مدل *FRVSR* توانسته نتایج بهتری را از خود برجای گذارد. مدل *SRGAN* از سایر مدل ها با تعداد دوره ۱۰۰، *PSNR* و *SSIM* پایین تری داشته است.

۵ - ۸ - نتایج آزمایش

در این بخش به آزمایش مدل های آموزش داده شده می پردازیم. جهت نیل به این هدف از مجموعه داده *Vid4* استفاده می کنیم که هر کدام شامل مجموعه ای از فریم های ویدیو به نام های *Calendar, City, Foliage, Walk* می باشند. نتایج آزمایش بر مبنای افزایش مقیاس ۴ برابری ویدیوها انجام شده است.



شکل ۱۷-۵-مجموعه داده *Vid4*

۵ - ۸ - ۱ - مقایسه مدل ها بر اساس دو معیار ***PSNR*** و ***SSIM***

همانند بخش ارزیابی جهت مقایسه عادلانه بین مدل ها، مدل *FRVSR* را به دو بخش *FRVSR10* و *FRVSR20* تقسیم بندی کردیم که بتوان مقایسه بهتری با *FRGANVSR* داشت.

جدول ۱۰-۵-نتایج *PSNR* در آزمایش مجموعه داده *Vid4*

Model/Dataset	Calendar	City	Foliage	Walk	Average
SRGAN	*28.8777*	*28.3725*	*28.5501*	*29.1216*	*28.73.05*
FRVSR10	*29.0981*	*29.5717*	*28.9026*	*30.4557*	*29.5070*
FRVSR20	*29.1009*	*29.6453*	*28.9247*	*30.5609*	*29.5579*
FRGANVSR	*29.1311*	*29.6152*	*28.9264*	*30.5973*	*29.5675*

جدول ۱۱-۵ - نتایج *SSIM* در آزمایش مجموعه داده *Vid4*

Model/Dataset	Calendar	City	Foliage	Walk	Average
SRGAN	*0.5853*	*0.5855*	*0.6098*	*0.7683*	*0.6372*
FRVSR10	*0.5736*	*0.5920*	*0.6275*	*0.7643*	*0.6393*
FRVSR20	*0.5748*	*0.5931*	*0.6293*	*0.7657*	*0.6407*
FRGANVSR	*0.5858*	*0.5825*	*0.6375*	*0.7678*	*0.6434*

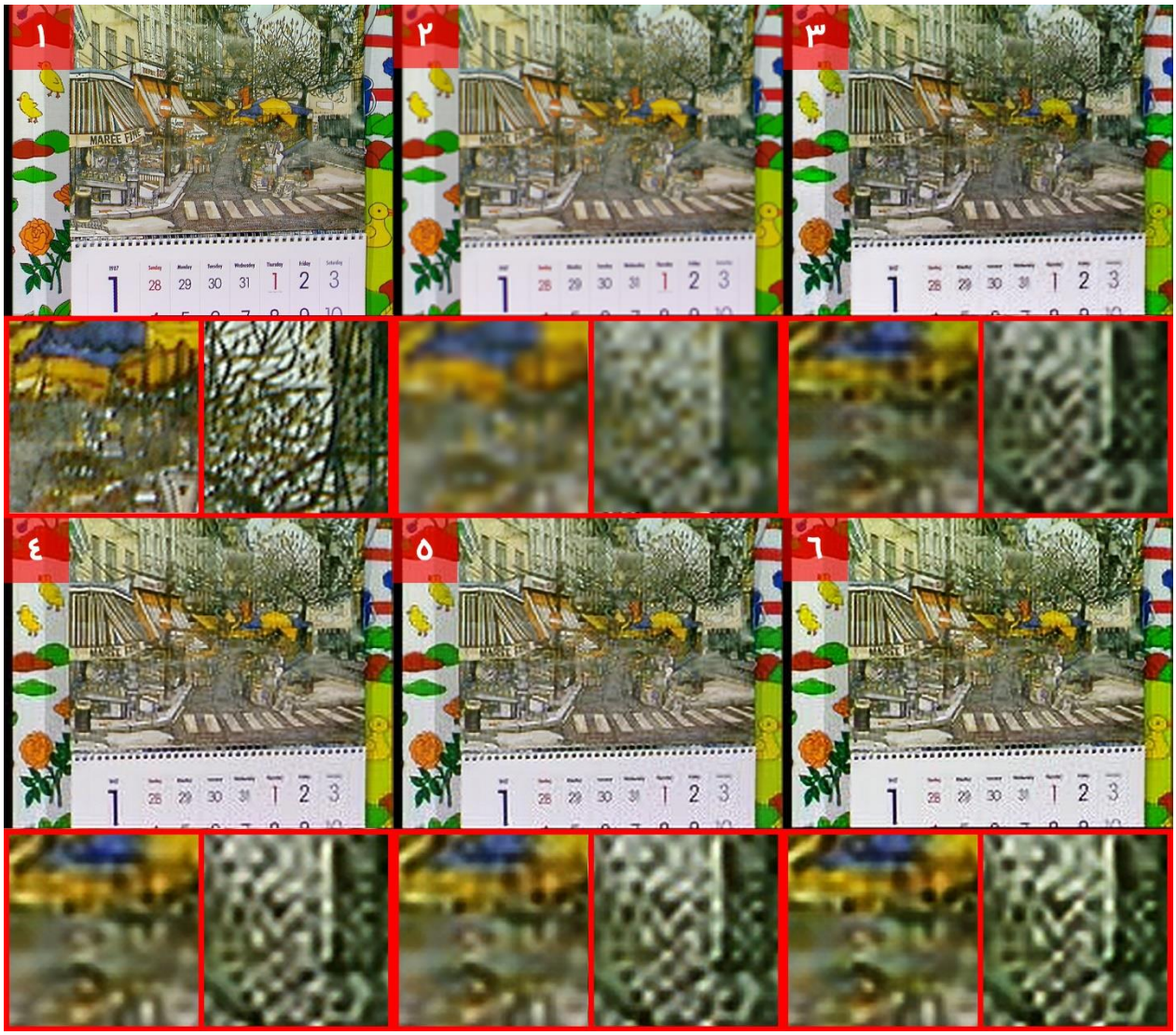
مدل پیشنهادی بر اساس هر دو معیار *PSNR* و *SSIM* عملکرد بهتری نسبت به سایر مدل ها داشته است.

نکته جالب توجه اینست که مدل پیشنهادی، نه تنها بر مدل *FRVSR* در تعداد دوره برابر غلبه کرده است، بلکه عملکرد بهتری نسبت به مدل *FRVSR* با ۲۰ دوره آموزش داشته است.

۵ - ۸ - ۲ - کیفیت بصری

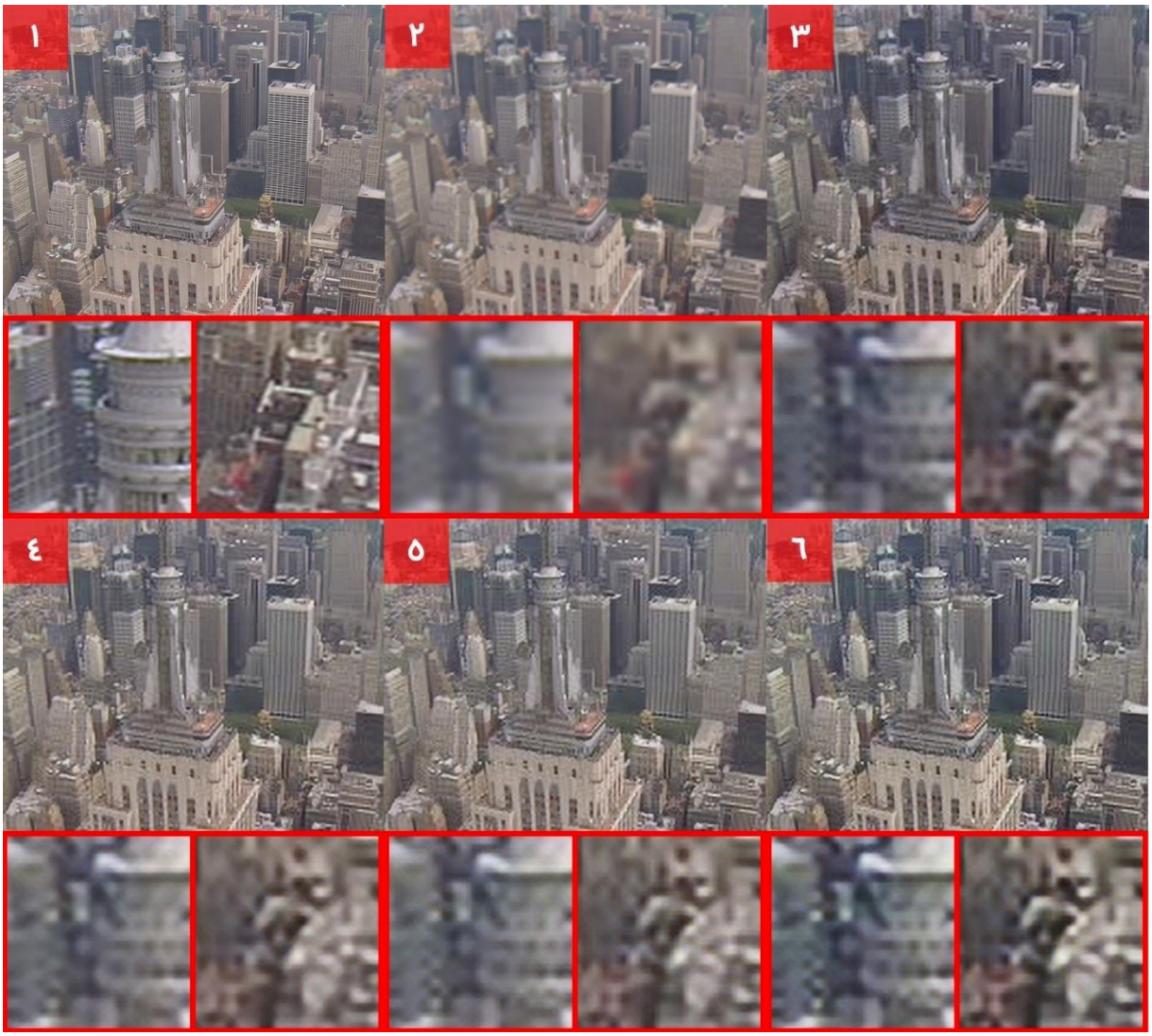
در این آزمایش عملکرد روش پیشنهادی را بر مبنای کیفیت بصری مورد آزمایش قرار می دهیم. به دلیل تشابه فراوان مدل های مورد مقایسه، کیفیت بصری نتایج در افزایش مقیاس ۴ برابری توسط چشم دشوار بوده و بدین منظور برخی بافت های تصاویر خروجی جهت مقایسه بزرگنمایی شدند. در این آزمایش تصویر اصلی (*Ground Truth*)، تصویر بزرگنمایی شده به روش درون یابی دوخطی (*Bilinear*)، روش *SRGAN*، روش *FRVSR10*، روش *FRVSR20* و روش پیشنهادی *FRGANVSR* مقایسه شده اند.

همانطور که در تصاویر پیداست، به دلیل الگوریتم های مشابه به کار رفته در مدل ها کیفیت تصاویر به سختی قابل تفکیک نسبت به یکدیگر هستند اما به طور کلی تصاویر فراتفکیک پذیر شده به روش پیشنهادی (*FRGANVSR*) از جزئیات و تباین بهتری برخوردار هستند.



شکل ۱۸-۵-مقایسه بصری مدل ها برای داده *Calendar*

تصویر (۱) - تصویر وضوح بالا (*Ground Truth*)، تصویر (۲)- درون یابی دو خطی، تصویر (۳)-مدل *SRGAN*، تصویر (۴)-مدل *FRVSR10*، تصویر (۵)-مدل *FRVSR20*، تصویر (۶)-مدل *FRVSRGAN* *(مدل پیشنهادی)*

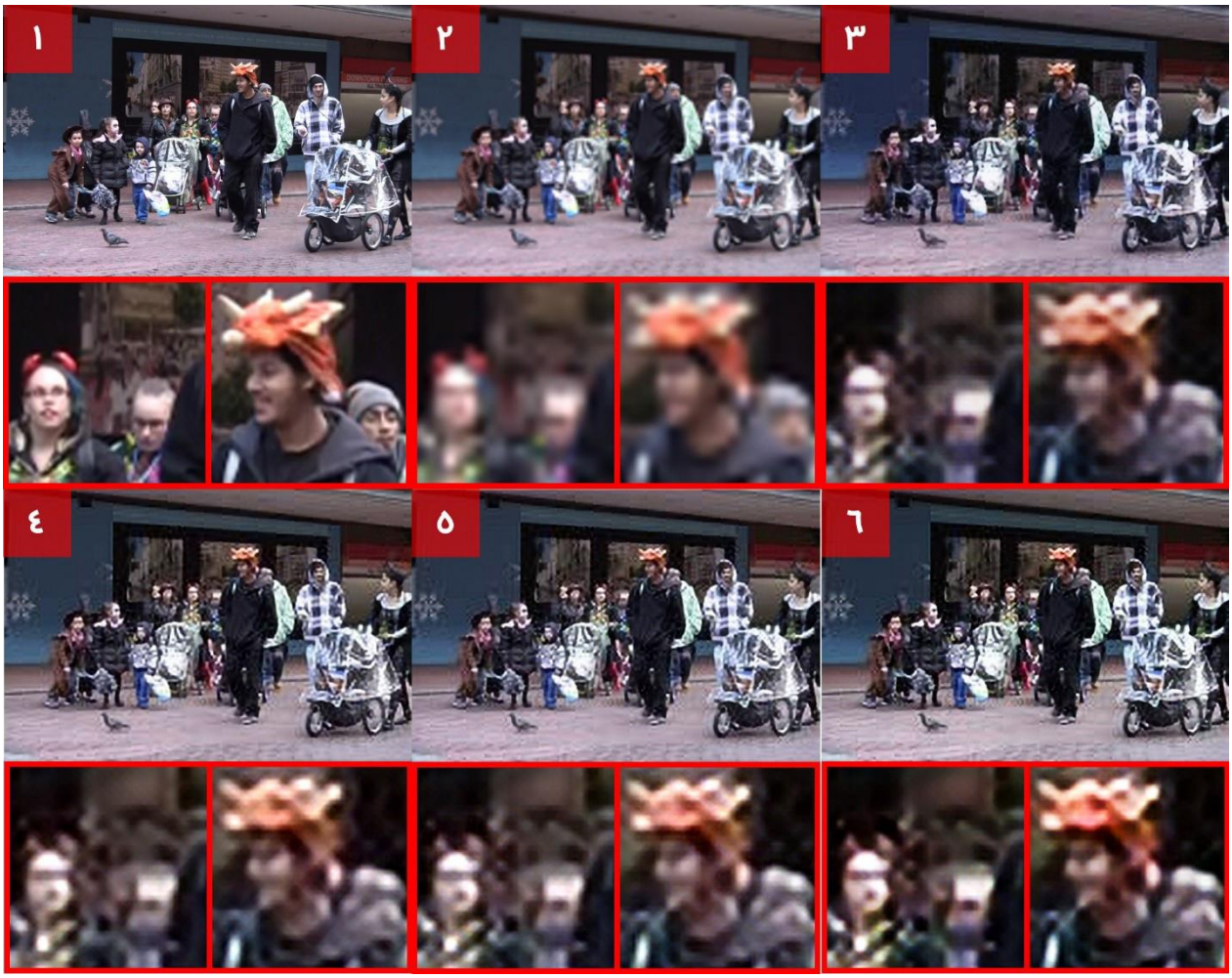


شکل ۱۹-۵- مقایسه بصری مدل ها برای داده *City*

تصویر (۱) - تصویر وضوح بالا (*Ground Truth*)، تصویر (۲)- درون یابی دو خطی، تصویر (۳)-مدل *SRGAN*، تصویر (۴)-مدل *FRVSR10*، تصویر (۵)-مدل *FRVSR20*، تصویر (۶)-مدل *FRVSRGAN* *(مدل پیشنهادی)*

شکل ۲۰-۵- مقایسه بصری مدل ها برای داده *Foliage*

تصویر (۱) - تصویر وضوح بالا (*Ground Truth*)، تصویر (۲)- درون یابی دو خطی، تصویر (۳)-مدل *SRGAN*، تصویر (۴)-مدل *FRVSR10*، تصویر (۵)-مدل *FRVSR20*، تصویر (۶)-مدل *FRVSRGAN* *(مدل پیشنهادی)*



شکل ۲۱-۵- مقایسه بصری مدل ها برای داده *Walk*

تصویر (۱) - تصویر وضوح بالا (*Ground Truth*)، تصویر (۲)- درون یابی دو خطی، تصویر (۳)-مدل *SRGAN*،
تصویر (۴)-مدل *FRVSR10*، تصویر (۵)-مدل *FRVSR20*، تصویر (۶)-مدل *FRVSRGAN* *(مدل پیشنهادی)*

۵ - ۸ - ۳ - مشخصه زمانی

عملکرد روش پیشنهادی را می توان از دیدگاه پیوستگی زمانی بین فریم های ویدیو نیز مورد بررسی قرار داد. بدین منظور تعداد مشخصی از فریم ها را به صورت پیوسته با برش مشخصی به اندازه یک پیکسل از عرض (یا ارتفاع) به صورت افقی (یا عمودی) در کنار هم قرار داده تا بتوان پیوستگی فریم ها را از دیدگاه بصری نیز اندازه گرفت. نتایج برای مدل های *GAN*، مدل *FRVSR10*، مدل پیشنهادی (*FRGANVSR*) و تصویر وضوح بالا (*Ground Truth*) در شکل ۲۶-۴ رسم شده است. همانطور که می بینیم روش پیشنهادی (*FRGANVSR*) همچنان دچار ناپیوستگی هایی می باشد. از دیدگاه بصری پیوستگی زمانی روش *FRVSR* نسبت به سایر روش ها عملکرد بهتری را از خود بر جای گذاشته اما روش پیشنهادی *FRGANVSR* نسبت به روش *SRGAN* عملکرد بهتری را در خصوص مشخصه پیوستگی زمانی نشان می دهد. به طور کلی روش پیشنهادی توانسته تعدیلی بین حفظ پیوستگی زمانی و کیفیت بصری مناسب ارائه نماید.

شکل ۲۲-۵- مقایسه مشخصه زمانی در بین ۴ مدل مختلف برای داده *Walk*

۵ – ۹ – بررسی مدل پیشنهادی در یک فریم ویدیویی

در این بخش به بررسی مدل پیشنهادی در یک فریم ویدیویی می پردازیم. مدل پیشنهادی در مقیاس ۴ برابر بررسی شده است. جهت این بررسی قصد داریم که یک ویدیوی با اندازه ۳۲۰x۱۸۰ را با افزایش مقیاس ۴ برابری با مدل پیشنهادی فراتفکیک پذیری نموده و آن را به یک ویدیوی *HD* با اندازه ۱۲۸۰x۷۲۰ تبدیل کنیم.

همچنین ویدیوی اصلی نیز در کیفیت *HD* یعنی ۱۲۸۰x۷۲۰ تهیه شده است تا به عنوان فریم مرجع (*Ground Truth*) از آن استفاده گردد. جهت مقایسه بهتر در کنار فراتفکیک پذیری انجام شده، فریم با اندازه ۳۲۰x۱۸۰ را با استفاده از درون یابی دومکعبی تا ۴ برابر بزرگ گرده ایم.

نتایج برای سه فریم از ویدیو به عنوان نمونه و برای سه حالت فریم اصلی (*Ground Truth*)، فریم درون یابی شده (*Bicubic Interpolation*) و فریم فراتفکیک پذیر شده با روش پیشنهادی (*FRGAN-VSR*) در ادامه آورده شده است.

همانطور که در نتایج مشخص است، روش پیشنهادی تا حد بسیار خوبی توانسته جزئیات فرکانس بالای تصاویر را بازیابی نموده و بافت تصاویر را تا حد امکان حفظ نماید.

از روش پیشنهادی می توان جهت تولید محتوای *4K* (۴۰۹۶x۲۱۶۰) از روی یک ویدیو *HD* نیز اقدام نمود.

شکل ۲۳-۵- فریم نمونه ۱ - مقایسه فریم وضوح بالا (۱) - روش پیشنهادی (۲) و روش درون یابی دومکعبی (۳)

شکل ۲۴-۵- فریم نمونه ۲- مقایسه فریم وضوح بالا (۱) - روش پیشنهادی (۲) و روش درون یابی دومکعبی (۳)

شکل ۲۵-۵- فریم نمونه ۳ -مقایسه فریم وضوح بالا (۱) - روش پیشنهادی (۲) و روش درون یابی دومکعبی (۳)

شکل ۲۶۵-۵- فریم نمونه ۴ -مقایسه فریم وضوح بالا (۱) - روش پیشنهادی (۲) و روش درون یابی دومکعبی (۳)

شکل ۲۷-۵-فریم نمونه ۵ -مقایسه فریم وضوح بالا (۱) - روش پیشنهادی (۲) و روش درون یابی دومکعبی (۳)

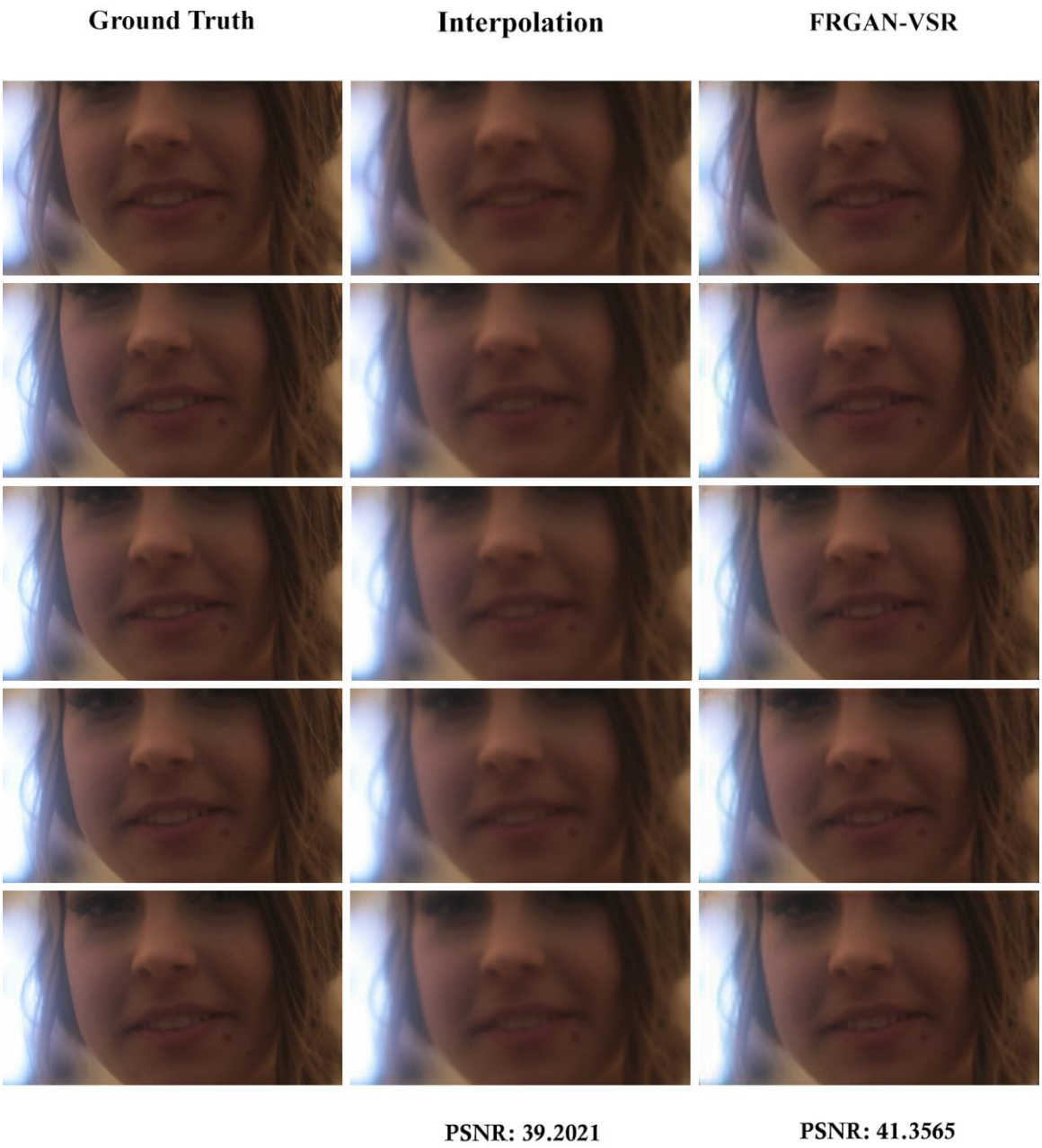
۵ – ۱۰ – بررسی مدل پیشنهادی در یک ویدیوی چندفریمی

در این آزمایش مدل پیشنهادی را روی یک ویدیو اعمال نمودیم. جهت نمایش نتایج در پایان نامه به نمایش ۵ فریم از هر ویدیو اکتفا نمودیم. نتایج این آزمایش برای ۵ ویدیوی ۵ فریمی ارائه شده است.

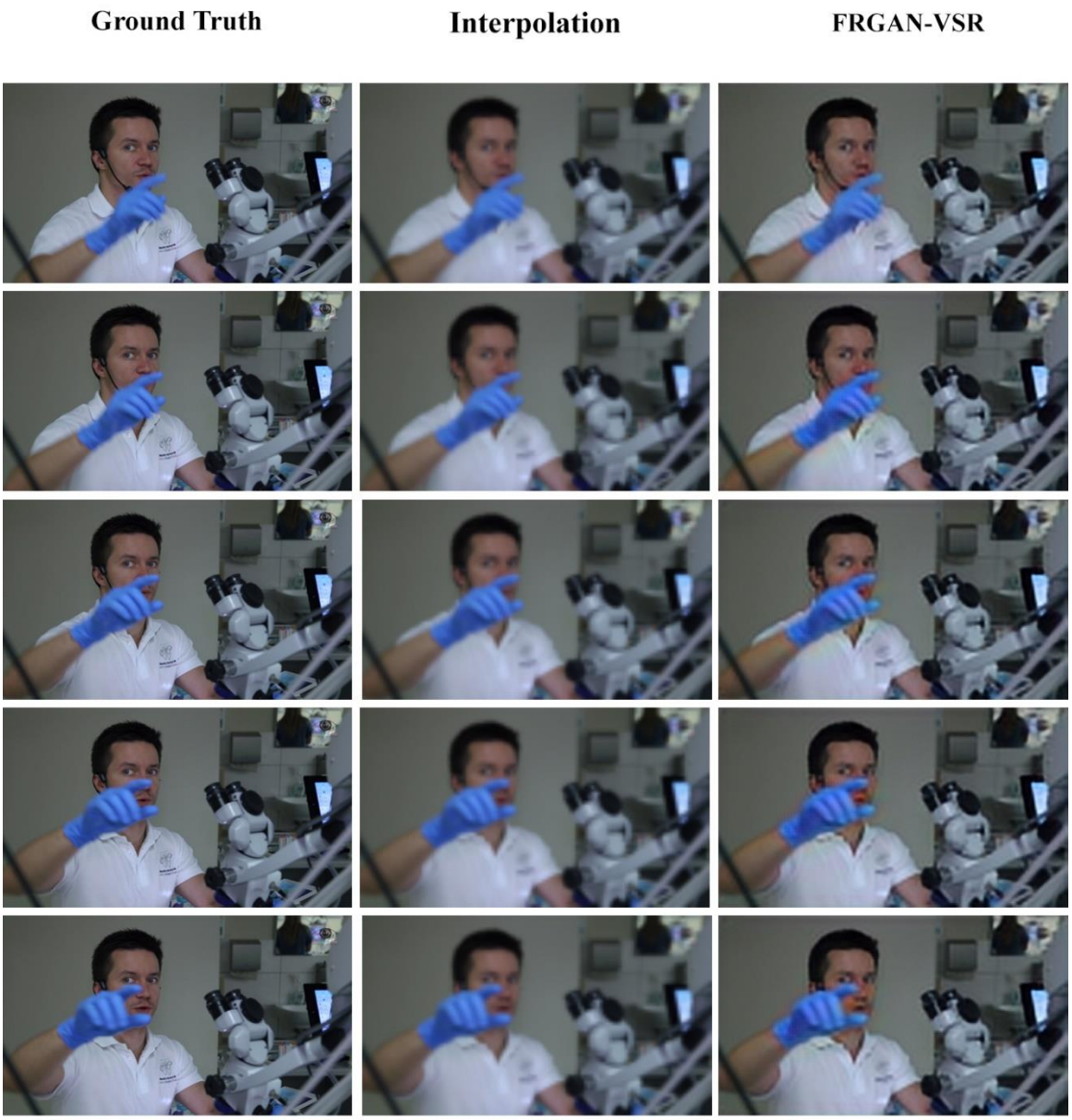
هر ویدیو دارای ابعاد ۱۱۲x۶۴ پیکسل بوده که با بزرگنمایی ۴ برابری آن را به ۴۴۸x۲۵۶ تبدیل نمودیم. همانند بخش ۴-۹ جهت مقایسه بهتر، فراتفکیک پذیری انجام شده به روش پیشنهادی، با ۵ فریم از ویدیوی اصلی (*Ground Truth*) و بزرگنمایی به روش درون یابی دومکعبی مقایسه شده است.

در این آزمایش علاوه بر مقایسه بصری انجام شده، روش پیشنهادی را با درون یابی دومکعبی با استفاده از معیار *PSNR* مقایسه نموده ایم. جهت نیل به این هدف، *PSNR* ویدیو به شکل فریم به فریم نسبت به ویدیوی اصلی اندازه گیری شده است. در نهایت *PSNR* کل ویدیو بر اساس میانگین گیری از *PSNR* های به دست آمده از تک تک فریم ها به دست آمده است. *PSNR* به دست آمده به ازای هر ویدیو در زیر هر شکل آورده شده است. همانطور که می بینیم، روش پیشنهادی علاوه بر وضوح بصری بالاتر نسبت به روش درون یابی، از *PSNR* بالاتری در کلیه ویدیوهای مقایسه شده نسبت به روش درون یابی برخوردار است. علاوه بر این جزئیات فرکانس بالا و بافت تصاویر در روش پیشنهادی حفظ شده و برخلاف درون یابی، با بزرگنمایی ۴ برابری از تاری تصویر جلوگیری شده است.

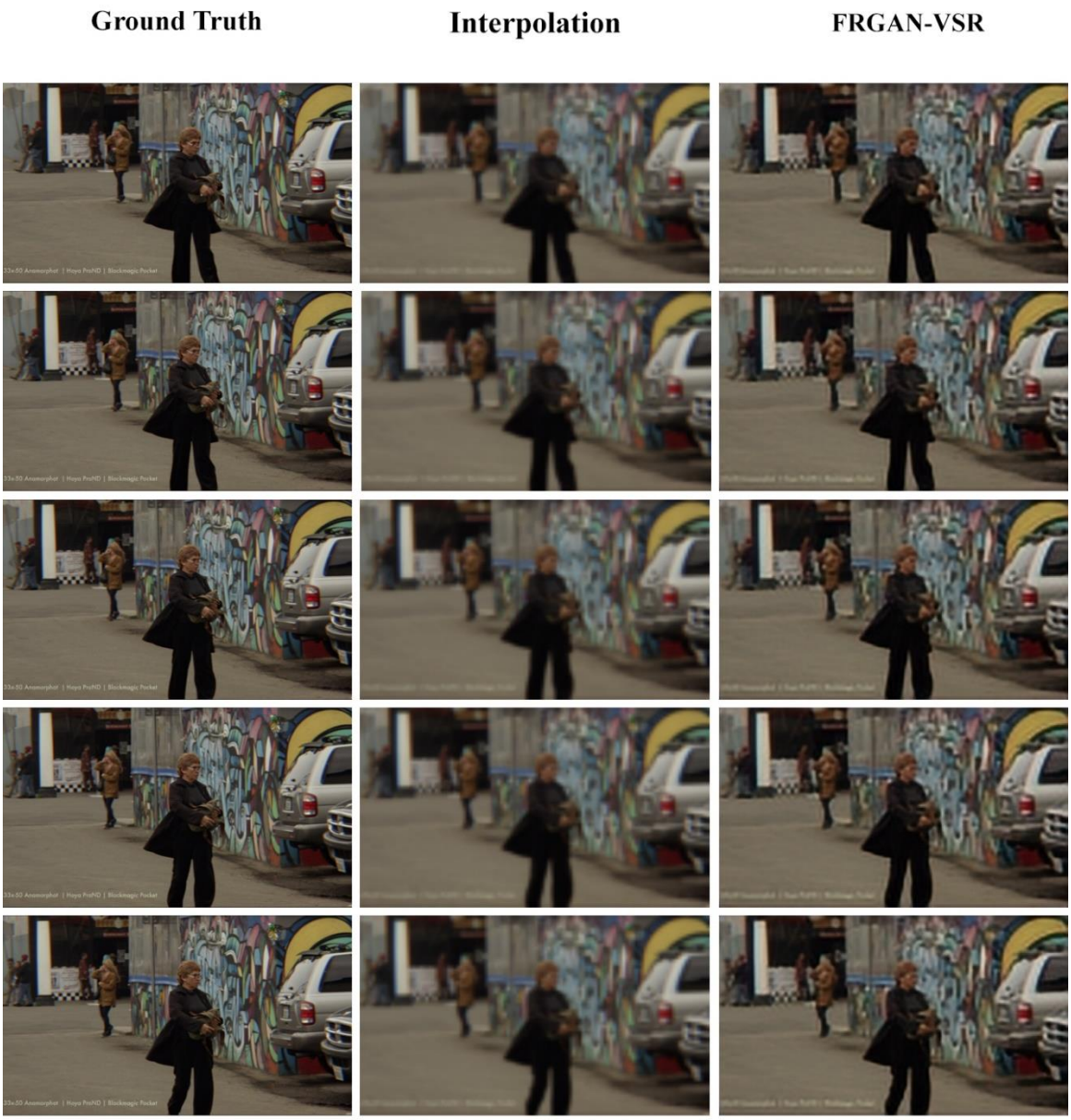
شکل ۲۸-۵-ویدیو نمونه ۱ - مقایسه PSNR روش پیشنهادی ، روش درون یابی دومکعبی و فریم اصلی



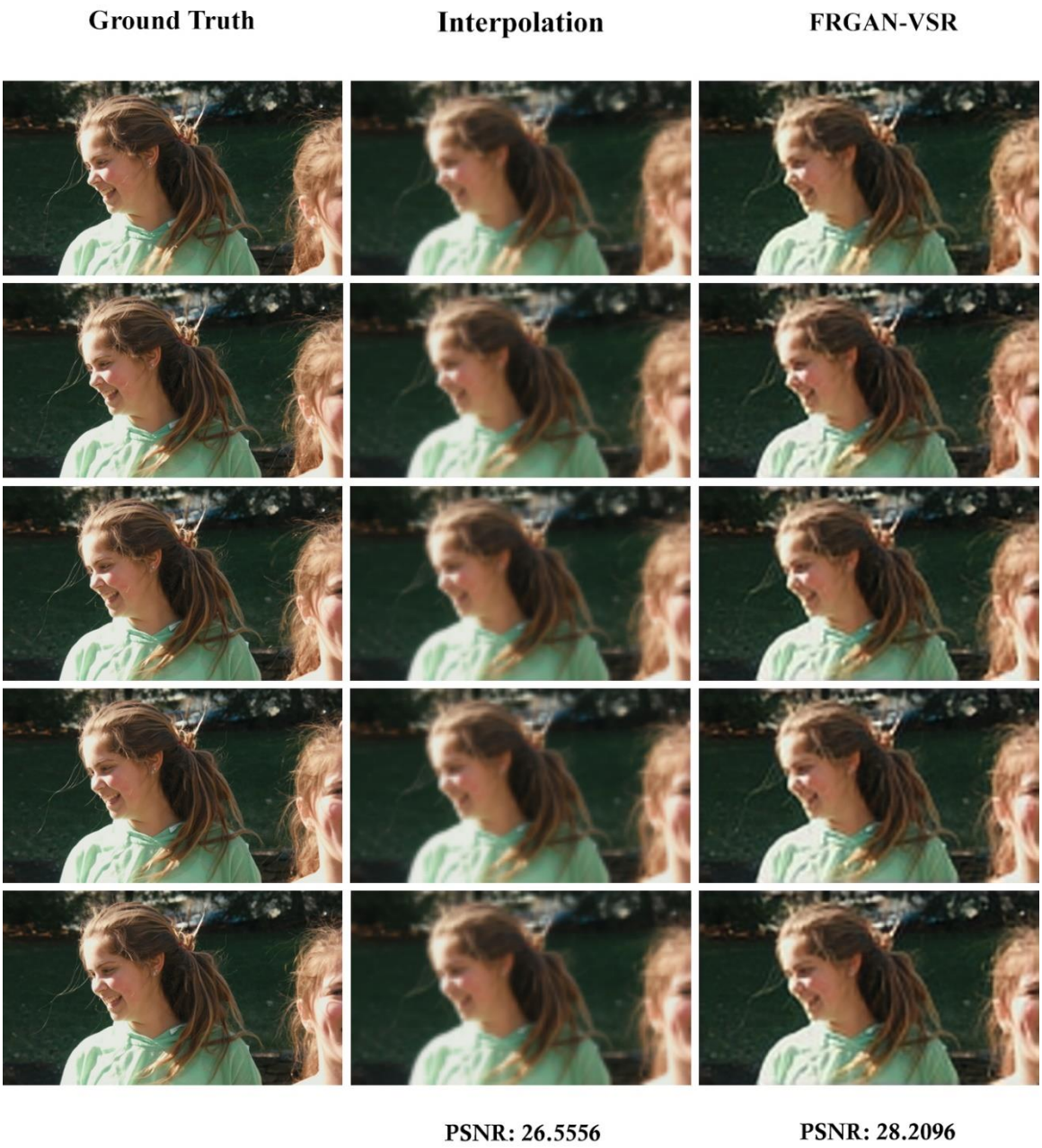
شکل ۲۹-۵-ویدیو نمونه ۲ - مقایسه *PSNR* روش پیشنهادی ، روش درون یابی دومکعبی و فریم اصلی



شکل ۳۰-۵-ویدیو نمونه ۳ - مقایسه *PSNR* روش پیشنهادی ، روش درون یابی دومکعبی و فریم اصلی



شکل ۳۱-۵-ویدیو نمونه ۴ - مقایسه *PSNR* روش پیشنهادی ، روش درون یابی دومکعبی و فریم اصلی



شکل ۳۲-۵-ویدیو نمونه ۵ - مقایسه *PSNR* روش پیشنهادی ، روش درون یابی دومکعبی و فریم اصلی

فصل ۶
نتیجه گیری و پژوهش های آینده

۶ – ۱ – ملاحظات

در این پژوهش به معرفی روشی پرداختیم که با استفاده از جدیدترین روش های مورد استفاده در الگوریتم های فراتفکیک پذیری بتواند به فراتفکیک پذیری ویدیو بپردازد. نگارنده در مطالعه و توسعه مدل پیشنهادی موارد ذیل را در نظر داشته است:

۶ – ۱ – ۱ - استفاده از شبکه های عمیق

همانطور که در فصل ۲ بیان شد، الگوریتم ها و رویکردهای گوناگونی جهت حل مسئله فراتفکیک پذیری مورد استفاده قرار گرفته است. در پژوهش حال حاضر سعی شد که از جدیدترین الگوریتم های شبکه های عمیق از جمله شبکه های عمیق پیچشی و شبکه های مولد تخاصمی استفاده گردد.

۶ – ۱ – ۲ - استفاده از توابع هزینه کارآمد

با توجه به کارایی ضعیف تابع هزینه مجموع مربع خطا در مدل فراتفکیک پذیر، سعی کردیم از ترکیب چند تابع هزینه برای دستیابی به عملکرد بهتر استفاده کنیم. بدین منظور به جای استفاده از مجموع مربع خطا به عنوان تنها تابع هزینه، از مجموع وزن دار چندین تابع هزینه در مدل استفاده کردیم.

۶ – ۱ – ۳ - مدلی مبتنی بر فراتفکیک پذیری ویدیو

فراتفکیک پذیری **ویدیو** یکی از فاکتورهای اساسی در توسعه مدل پیشنهادی بوده است. نگارنده با علم به این موضوع به دنبال رویکردی بود که علاوه بر کیفیت مطلوب در فراتفکیک پذیری تصویر، از اطلاعات موجود در فریم های موجود دیگر نیز استفاده نماید. بدین منظور از یک شبکه آموزش پذیر مبتنی بر شار نوری استفاده شد که بتوان اطلاعات فریم های قبل را با فریم کنونی تلفیق نمود و نتایج پیوسته بهتری را در ویدیو کسب نمود.

۶ - ۱ - ۳ - ۱ - عملکرد مطلوب در بزرگ نمایی ۴ برابر

این پژوهش منطبق بر عدم تناسب پیشرفت نمایشگرهای مدرن و محتوای متناسب با این گونه نمایشگرها بنا نهاده شده است. به بیان دیگر نمایشگرهای حال حاضر توانایی نمایش محتواهایی با وضوح بسیار بالا نظیر *4K* و *8K* را دارند. در صورتی که اکثر محتوای تولیدی موجود منطبق با نمایشگرهایی با وضوح بسیار پایین تر می باشند. لذا بزرگنمایی ۲ برابر به حل این موضوع کمکی نکرده و نیاز به پیشنهاد روشی بود که قادر به بزرگنمایی ۴ تا ۸ برابر با کیفیت مطلوب باشد. نگارنده با علم به این موضوع از شبکه های مولد تخاصمی (*GAN*) که از آن می توان به عنوان یکی از ایده های انقلابی دهه اخیر در یادگیری ماشین نام برد استفاده کرده است که توانایی بازسازی تصاویر با بزرگنمایی ۴ تا ۸ برابری را با کیفیت مطلوب و قابل توجه دارا می باشند.

۶ – ۲ – نتیجه گیری

روش پیشنهادی این پایان نامه نه بر اساس بهبود عملکرد یک مدل و بلکه مبتنی بر معرفی یک چهارچوب گسترده بود که علاوه بر قابلیت استفاده در کاربردهای گوناگون، می تواند به عنوان مرجعی جهت مطالعات آتی مورد ملاحظه قرار گیرد. مدل پیشنهادی دربردارنده سه شبکه ی مبتنی بر یادگیری عمیق و محاسبه گر شار نوری بوده و با رویکرد شبکه های مولد تخاصمی ارائه گردید و توانست به عنوان یکی از مدل های مدرن، نتایج قابل توجهی را ارائه دهد. هرچند مدل پیشنهادی علی رغم کیفیت قابل توجه و عملکرد قابل قبول، همچنان از پیچیدگی محاسباتی رنج می برد و جزو مدل های مبتنی بر یادگیری با محاسبات بالا و پردازش سنگین قلمداد می گردد.

۶ - ۳ - پژوهش های آینده

همانطور که در بخش ۵-۲ بیان شد، مدل پیشنهادی مبتنی بر معرفی یک چهارچوب گسترده بود که پارامترهای فراوانی در عملکرد آن موثرند. در این بخش به معرفی ایده هایی جهت بهبود مدل پیشنهادی می پردازیم که می توانند در مطالعات آتی لحاظ گردند.

۶ - ۳ - ۱ - توسعه مجموعه داده

روش پیشنهادی صرفاً بر اساس مجموعه داده *Vimeo_toflow* آموزش دیده که البته از بهترین مجموعه داده های قابل استفاده در آموزش شبکه های فراتفکیک پذیر ویدیو می باشد. جهت بهبود مدل و یادگیری بهتر شبکه می توان از مجموعه داده های متفاوتی نظیر *Holopix50k* [۵۸] که شامل ۵۰.۰۰۰ تصویر و جزو جدیدترین و بزرگترین مجموعه های داده موجود در فراتفکیک پذیری می باشد، استفاده نمود.

۶ - ۳ - ۲ - افزایش طول دوره آموزش شبکه ها

مدل پیشنهادی در ۱۰ دوره آموزش دیده و با این حال نتایج قابل قبولی را ارائه می دهد. می توان آن را برای دوره های بیشتری جهت افزایش کیفیت آموزش داد. نتایج به دست آمده تا ۱۰ دوره، افزایش کیفیت مدل پیشنهادی را با افزایش طول دوره ها بسیار محتمل می داند.

۶ - ۳ - ۳ - بهینه سازی توابع هزینه

هرچند مدل پیشنهادی جهت بهبود کیفیت خود از توابع هزینه گوناگون به جای مجموع مربعات خطا استفاده کرده است. اما از سوی دیگر این ایده، باری اضافی را به پردازش های مدل تحمیل کرده است. مطالعه بر روی توابع هزینه بهینه با حفظ کیفیت مدل و کاهش بار محاسباتی می تواند مبنایی جهت تحقیقات آتی باشند.

۶ - ۳ - ۴ - بهینه سازی فراپارامترها

تعدد فراپارامترهای مورد استفاده در مدل پیشنهادی خصوصاً ضرایب مورد استفاده در توابع هزینه موجب شد که پژوهش حال حاضر به بررسی عمیق تر ضرایب توابع هزینه بهینه نپردازد. پژوهش های آتی می تواند مبتنی بر بهینه سازی ضرایب انتخابی توابع هزینه و سایر فراپارامترها باشند.

۶ - ۳ - ۵ - بهینه سازی شبکه های آموزش پذیر

تعدد شبکه های آموزش پذیر، پژوهشگر را بر آن می دارد که به بهینه سازی شبکه های آموزش پذیر پرداخته و با تغییر معماری شبکه ها به ارتقا کیفیت مدل پیشنهادی بپردازد. بده و بستان[1] های صورت گرفته در خصوص کیفیت مدل و حجم پردازش های محاسباتی از جمله مواردی ست که می تواند به شکل گسترده ای در شبکه های مدل پیشنهادی، مورد مطالعه قرار گیرد.

۶ - ۳ - ۶ - سایر

علاوه بر ملاحظات فوق جهت بهبود مدل پیشنهادی می توان به بهبود چهارچوب الگوریتم نیز اقدام نمود.

اشیای جلوی صحنه[2] توجه بیشتری را نسبت به تصاویر پس زمینه به خود جلب می کنند و برای یک ناظر انسانی از اهمیت بالاتری برخوردار است. جهت بهبود این کیفیت بصری می توان شیوه الگوریتم را به نوعی تغییر داد که به جای یک فریم از چندین فریم جهت دریافت جزئیات اشیا اقدام نماید و برای سایر موارد موجود در صحنه از یک فریم استفاده کند. در این رویکرد هم دغدغه پردازش اضافی مد نظر قرار می گیرد و هم از جزئیات بهتری در خصوص اشیای جلوی صحنه برخوردار خواهیم شد.

[1] Trade-off
[2] Foreground

مراجع

[1] Park, S. C, Park, M. K, & Kang, M. G, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine,* vol. 20, no. 3, pp. 21-36, 2003.

[2] Dong, C., Loy, C. C., He, K., & Tang, X, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE,* vol. 38, no. 2, pp. 295-307, Feb 2016.

[3] Kouame, D., & Ploquin, M, "Super-resolution in medical imaging : An illustrative approach through ultrasound," *IEEE International Symposium on Biomedical Imaging: From Nano to Macro,* 2009.

[4] Hasan, D and Gholamreza, A,., "Discrete Wavelet Transform-Based Satellite Image Resolution Enhancement," *IEEE,* vol. 49, no. 6, pp. 1997-2004, 2011.

[5] Zhang, H., Shen, H., & Li, P, "A super-resolution reconstruction algorithm for surveillance images," *Signal Processing,* vol. 90, no. 3, pp. 848-859, 2010.

[6] Singh, A., & Singh, J, "Super Resolution Applications in Modern Digital Image Processing," *International Journal of Computer Applications,* vol. 150, no. 2, pp. 6-8, September 2016.

[7] Müller, M. U., Ekhtiari, N., Almeida, R. M., & Rieke, C, "Super-resolution of multispectral satellite images using convolutional neural networks," *Computer Vision and Pattern Recognition,* vol. V, pp. 33-40, 2020.

[8] Wang, Z., Jiang, K., Yi, P., Han, Z., & He, Z, "Ultra-dense GAN for satellite imagery super-resolution," *Elsevier,* vol. 398, pp. 328-337, 2020.

[9] Gupta, R., Sharma, A., & Kumar, A, "Super-Resolution using GANs for Medical Imaging," *Procedia Computer Science,* vol. 173, pp. 28-35, 2020.

[10] Mahapatra, D., Bozorgtabar, B., & Garnavi, R, "Image super-resolution using progressive generative adversarial networks for medical image analysis," *Elsevier,* vol. 7, pp. 30-39, 2019.

[11] Durisic, N., Cuervo, L. L., & Lakadamyali, M, "Quantitative super-resolution microscopy: pitfalls and strategies for image analysis," *Current Opinion in Chemical Biology,* vol. 20, pp. 22-28, 2014.

[12] Godin, A. G., Lounis, B., & Cognet, L, "Super-resolution Microscopy Approaches for Live Cell Imaging," *Biophysical Journal*, vol. 107, no. 8, pp. 1777-1784, 2014.

[13] GUO, R., SHI, X., ZHU, Y., & YU, T, "Super-resolution reconstruction of astronomical images using time-scale adaptive normalized convolution," *Chinese Journal of Aeronautics*, vol. 31, no. 8, pp. 1752-1763, 2018.

[14] Romano, Y., Isidoro, J., & Milanfar, P, "RAISR: Rapid and Accurate Image Super Resolution," *IEEE*, vol. 3, no. 1, pp. 110-125, 2016.

[15] W. Wendell, "Google RAISR | Image Resolution Enhancement Straight Out Of CSI," 2016. [Online]. Available: https://www.slrlounge.com/google-raisr-image-resolution-enhancement-straight-out-of-csi/. [Accessed 5 1 2021].

[16] Milanfar, P et al., "Enhance! RAISR Sharp Images with Machine Learning," Google, 14 November 2016. [Online]. Available: https://ai.googleblog.com/2016/11/enhance-raisr-sharp-images-with-machine.html.

[17] Younghyun Jo, Seoung Wug Oh, Jaeyeon Kang, Seon Joo Kim, "Deep Video Super-Resolution Network Using Dynamic Upsampling Filters Without Explicit Motion Compensation," *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018.

[18] Na, B., & Fox, G. C, "Object Detection by a Super-Resolution Method and a Convolutional Neural Networks," *IEEE International Conference on Big Data (Big Data)*, 2018.

[19] Chuang, C.-H., Tsai, L.-W., Deng, M.-S., Hsieh, J.-W., & Fan, K.-C, "Vehicle licence plate recognition using super-resolution technique," *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2014.

[20] Gong, Z., Yu, F., & Tang, Y, "Real-time Video Image Super-Resolution Network for Mobile Terminals," *3rd International Conference on Circuits, System and Simulation (ICCSS)*, 2019.

[21] Aubel, C., Stotz, D., & Bolcskei, H, "Super-resolution from short-time Fourier transform measurements," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.

[22] Ray Choudhury, A., & Dahake, V. R, "Image Super Resolution Using Fractional Discrete Wavelet Transformand Fractional Fast Fourier Transform," *IEEE 2nd International Conference on I-SMAC*, 2018.

[23] Nasrollahi, K., & Moeslund, T. B, "Super-resolution: a comprehensive survey," *Machine Vision and Applications,* vol. 25, no. 6, pp. 1423-1468, 2014.

[24] Irani, M., & Peleg, S, "Improving Resolution by Image Registration," *Graphical Models and Image Processing,* vol. 53, pp. 231-239, 1991.

[25] Peleg, M. Irani and S., "Image sequence enhancement using multiple motions analysis," *Proceedings of International Conference on Computer Vision and Pattern Recognition,* pp. 216-222, 1992.

[26] M. &. F. A. Elad, "Super-resolution reconstruction of continuous image sequences," *Proceedings of International Conference on Image Processing,* vol. 3, pp. 459-463, 1999.

[27] Elad, M., & Feuer, A, "Super-resolution restoration of an image sequence: adaptive filtering approach," *IEEE Transactions on Image Processing,* vol. 8, no. 3, pp. 387-395, 1999.

[28] Farsiu, S., Robinson, D., Elad, M., Milanfar, P, "Dynamic demosaicing and color Super-Resolution video sequences," *Proceedings of SPIE Conference on Image Reconstruction from Incomplete Data,* 2004.

[29] Farsiu, S., Robinson, D., Elad, M., Milanfar, P, "Fast and robust multi-frame super-resolution," vol. 13, pp. 1327-1344, 2004.

[30] Yang, C.-Y., Ma, C., & Yang, M.-H, "Single-Image Super-Resolution: A Benchmark," *Computer Vision – ECCV 2014,* pp. 372-386, 2014.

[31] R. Fattal, "Image upsampling via imposed edge statistics," *SIGGRAPH,* vol. 26, no. 3, p. 95, 2007.

[32] Sun, J., Sun, J., Xu, Z., Shum, H.Y, "Image super-resolution using gradient profile prior," *CVPR,* 2008.

[33] Shan Q., Li Z, Jia J, Tang C.K, "Fast image/video upsampling," *SIGGRAPH Asia,* pp. 1-7, 2008.

[34] Kim, K. I & Kwon, Y, "Single-image super-resolution using sparse regression and natural image prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 32, no. 6, pp. 1127-1133, 2010.

[35] Lee, J., Lee, J., & Yoo, H.-J, "SRNPU: An Energy-Efficient CNN-Based Super-Resolution Processor With Tile-Based Selective Super-Resolution in Mobile Devices," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems,* vol. 10, no. 3, pp. 320-334, 2020.

[36] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2017.

[37] Tai, Y.-W., Liu, S., Brown, M. S., & Lin, S, "Super Resolution using Edge Prior and Single Image Detail Synthesis," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 2010.

[38] Liu, H., Ruan, Z., Zhao, P et al, "Video Super Resolution Based on Deep Learning: A Comprehensive Survey," *Computer Vision and Pattern Recognition*, 2020.

[39] Chauhan, K., Patel, H., Dave, R., Bhatia, J, "Advances in Single Image Super-Resolution: A Deep Learning Perspective," *Proceedings of First International Conference on Computing, Communications, and Cyber-Security,* 2020.

[40] Sajjadi, M., Vemulapalli R., Brown, M., "Frame-Recurrent Video Super-Resolution," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

[41] Ying, X., Wang, L., Wang, Y., Sheng, W., An, W., Guo, Y., "Deformable 3D Convolution for Video Super-Resolution," *IEEE Signal Processing,* 2020.

[42] Lucas, A., Lopez-Tapia, S., Molina, R., & Katsaggelos, A. K., "Generative Adversarial Networks and Perceptual Losses for Video Super-Resolution," *IEEE Transactions on Image Processing,* vol. 28, no. 7, pp. 3312-3327, 2019.

[43] Chao, J. Guo and H., "Building an end-to-end spatial-temporal convolutional network for video super-resolution," *AAAI Conf Artif. Intell,* pp. 4053-4060, 2017.

[44] S. Li, F. He, B. Du, L. Zhang, Y. Xu, and D. Tao, "Fast spatio-temporal residual network for video super-resolution Pattern Recognit," *Proc. IEEE Conf. Comput Vis,* 2019.

[45] Yi, P., Wang, Z., Jiang, K., Jiang, J., & Ma, J, "Progressive fusion video super-resolution network via exploiting non-local spatio-temporal correlations," *IEEE Int. Conf. Comput. Vis,* pp. 3106-3115, 2019.

[46] Moghimi, M. K., & Mohanna, F, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," *IEEE Conference on Computer Vision and Pattern Recognition*, p. 1874–1883, 2016.

[47] He, K., Zhang, X., Ren, S., Sun, J., "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition,* p. 770–778, 2016.

[48] Kim, J., Lee, J. K., & Lee, K. M, "Deeply-recursive convolutional network for image super resolution," *IEEE Conference on Computer Vision and Pattern Recognition,* p. 1637–1645, 2016.

[49] Tai, Y., Yang, J., Liu, X, "Image super-resolution via deep recursive residual network," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* p. 3147–3155, 2017.

[50] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., "Densely connected convolutional networks," *IEEE Conference on Computer Vision and Pattern Recognition,* p. 4700–4708, 2017.

[51] Tong, T., Li, G., Liu, X., & Gao, Q, "Image super-resolution using dense skip connections," *IEEE International Conference on Computer Vision,* p. 4799–4807, 2017.

[52] Johnson, J et al., "Perceptual losses for real-time style transfer and super-resolution," *European Conference on Computer Vision,* p. 694–711, 2016.

[53] Horn, B.K.P and Schunck, "Determining optical flow," *Artificial Intelligence,* pp. 185-203, 1981.

[54] Hur, J & Roth. S, "Optical Flow Estimation in the Deep Learning Age," *Department of Computer Science,* 2020.

[55] Jaderberg, M., Simonyan, K., Zisserman, A and K. Kavukcuoglu, "Spatial transformer networks," *NIPS,* 2015.

[56] Agustsson, E., & Timofte, R, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," 2017.

[57] Xue, T et al, "Video Enhancement with Task-Oriented Flow," *International Journal of Computer Vision (IJCV),* vol. 127, 2019.

[58] A. Limarc, "Introducing theHolopix50k Dataset for Image Super-Resolution," 6 9 2020. [Online]. Available: https://hackernoon.com/introducing-theholopix50k-dataset-for-image-super-resolution-l95k3unm. [Accessed 5 1 2021].

[59] M. Elad, A. Feuer, "Super-resolution reconstruction," *Proceedings of International Conference on Image Processing,* vol. 3, pp. 459-463, 1999.

[60] Li, C., Wand, M., "Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 2479–2486, 2016.

[61] Goodfellow, I, "Generative Adversarial Networks," 2014.

[62] Haris, M., Shakhnarovich, G., & Ukita, N, "Deep Back-Projection Networks for Single Image Super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[63] Liu, H., Ruan, Z., Zhao, P et al, "Video Super Resolution Based on Deep Learning: A Comprehensive Survey," *Computer Vision and Pattern Recognition*, 2020.

Abstract

Nowadays, having high quality images and videos is undoubtedly of a matter of importance. Super-Resolution technique is a software approach within the scope of signal processing, employed for enhancing the quality of an image by increasing the number of pixels along with decreasing the noise. In this technique, a high-resolution image will be created from one or multiple low-resolution images.In this thesis, we will introduce a video super-resolution method founded on learning approach and based on a convolutional neural network, in which a trainable Optical Flow network is used for gathering the information between the current frame and the previous one. The extra information received from the Optical Flow network, will be added to the current low-resolution frame. Afterward, this data will be sent to the Super-Resolution block where the current frame is super-resolved by a deep neural network. finally, the reconstructed high-resolution frame as a generator block will be given to a discriminator block in a Generative Adversarial Network (GAN) framework which will finally result in presenting a Photo-Realistic Super-Resolved frame.

The suggested approach, taken through using three trainable blocks, employing convolutional neural networks, receiving data from Optical Flow and applying Generative Adversarial Networks, has succeeded in magnifying the video 4 times without any conspicuous reduction in quality.

In the end, the proposed model (FRGAN-VSR) was compared with the SRGAN model (Omitting Flow Network from the proposed method) and also FRVSR (Removing Generative Adversarial Network from our model) and evaluated in the Training, Evaluating and Testing phases. To assess the visual quality, we tested our model based on PSNR and SSIM measurements. The proposed model has also been tested in a few sample frames along with a few multi-frame videos. Furthermore, we analyzed the temporal profile of the model and compared its temporal consistency with other models. The evaluations and comparison with previous methods demonstrate that the proposed model can outperform the current state of the art.

Keywords: Super-Resolution, Video, Resolution, Convolutional neural network, Optical Flow, GAN

Shahrood University of Technology

Faculty of Electrical Engineering

M.Sc Thesis in Communication Systems Engineering

Super-Resolution in HD Video by using Deep Networks

By: Ali Mozafari

Supervisor:

Dr. Hossein Khosravi

January 2021