



دانشگاه صنعتی امیرکبیر

(پلی‌تکنیک تهران)

دانشکده مهندسی کامپیووتر

پروژه تحقیقاتی درس بینایی ماشین

ترمیم تصویر با استفاده از یادگیری عمیق

نگارش

عطیه غفارلوی مقدم

استاد درس

دکتر رضا صفابخش

بهمن ۱۴۰۲

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ

چکیده

مساله‌ی ترمیم تصویر یکی از مسائل با سابقه‌ی طولانی در حوزه‌ی بینایی ماشین است که به وظیفه‌ی پرکردن نواحی از دست رفته‌ی تصویر با پیکسل‌هایی که هم از نظر معنایی و هم از نظر بافت با تصویر هماهنگی دارند، می‌پردازد. چالش‌های مطرح در این مساله را به طور کلی می‌توان در دو دسته قرار داد؛ گروه اول چالش‌هایی هستند که متوجه تولید بافت، رنگ و ویژگی‌های ظاهری با جزئیات سازگار با بخش‌های تخریب نشده هستند و گروه دوم چالش‌هایی هستند که متوجه تولید نواحی سازگار از لحاظ محتوایی و معنایی هستند. در این گزارش پنج روش ترمیم تصویر مطرح شده است که به چالش‌های بالا از جنبه‌های متفاوتی می‌نگرند و در صدد حل آنها برمی‌آیند.

واژه‌های کلیدی:

ترمیم تصویر، شبکه‌های مولد تقابلی، بازیابی تصویر، بینایی ماشین، پردازش تصویر، ساختار کدگذار-کدگشا

فهرست مطالب

| صفحه | عنوان | صفحه |
|------|---|------|
| ۱ | ۱ مقدمه | ۱ |
| ۵ | ۲ تبدیل‌های محتوایی تجمعی در شبکه‌های مولد تقابلی | ۲ |
| ۶ | ۱-۲ ایده‌ی اصلی | |
| ۷ | ۲-۲ معماری | |
| ۹ | ۳-۲ آموزش شبکه | |
| ۹ | ۴-۲ نتایج | |
| ۱۱ | ۳ شبکه‌ی تعمیر و شبکه‌ی بهینه سازی | ۳ |
| ۱۲ | ۱-۳ ایده‌ی اصلی | |
| ۱۲ | ۲-۳ معماری | |
| ۱۲ | ۱-۲-۳ شبکه‌ی تعمیر | |
| ۱۴ | ۲-۲-۳ شبکه‌ی بهینه سازی | |
| ۱۴ | ۳-۳ آموزش شبکه | |
| ۱۴ | ۱-۳-۳ توابع خطای شبکه‌ی تعمیر | |
| ۱۶ | ۲-۳-۳ توابع خطای شبکه‌ی بهینه سازی | |
| ۱۶ | ۴-۳ نتایج | |
| ۱۸ | ۴ ترمیم تصویر دومسیره | ۴ |
| ۱۹ | ۱-۴ ایده‌ی اصلی | |
| ۲۰ | ۲-۴ معماری | |
| ۲۰ | ۱-۲-۴ مسیر معکوس | |
| ۲۱ | ۲-۲-۴ مسیر روبه جلو | |
| ۲۲ | ۳-۴ آموزش شبکه | |
| ۲۲ | ۴-۴ نتایج | |
| ۲۴ | ۵ ترمیم تصویر با بهبود محلی و سراسری | ۵ |
| ۲۵ | ۱-۵ ایده‌ی اصلی | |
| ۲۶ | ۲-۵ معماری | |
| ۲۶ | ۱-۲-۵ شبکه‌ی ترمیم درشت مقیاس | |
| ۲۷ | ۲-۲-۵ شبکه‌ی بهبود محلی | |
| ۲۷ | ۳-۲-۵ شبکه‌ی بهبود سراسری مبتنی بر مکانیزم توجه | |
| ۲۷ | ۳-۵ آموزش شبکه | |

| | |
|----------|--------------------------------------|
| ۲۷ | ۱-۳-۵ توابع خطای شبکه‌ی درشت مقیاس |
| ۲۸ | ۲-۳-۵ توابع خطای شبکه‌ی بهبود محلی |
| ۲۹ | ۳-۳-۵ توابع خطای شبکه‌ی بهبود سراسری |
| ۲۹ | ۴-۵ نتایج |
| ۳۱ | ۶ ترمیم تصویر با جزئیات فرکانس بالا |
| ۳۲ | ۱-۶ ایده‌ی اصلی |
| ۳۲ | ۲-۶ معماری |
| ۳۳ | ۱-۲-۶ شبکه‌ی ترمیم درشت مقیاس |
| ۳۳ | ۲-۲-۶ شبکه‌ی سوپررزولوشن |
| ۳۴ | ۳-۲-۶ شبکه‌ی بهبود با رزلوشن بالا |
| ۳۴ | ۳-۶ نتایج |
| ۳۶ | کتابنامه |

فهرست تصاویر

صفحه

شکل

| | | |
|-----|--|----|
| ۱-۱ | انواعی از تخریب‌ها در مساله‌ی ترمیم تصویر | ۲ |
| ۱-۲ | معماری تمت با شبکه‌ی مولد تقابلی | ۷ |
| ۲-۲ | ساختار بلوک تبدیل محتوای تجمیعی | ۸ |
| ۳-۲ | نتایج ترمیم تصویر در روش تمت با شبکه‌ی مولد تقابلی و سایر شبکه‌ها | ۱۰ |
| ۱-۳ | معماری شبکه‌ی تعمیر در مدل آرنون | ۱۳ |
| ۲-۳ | معماری شبکه‌ی بهینه‌سازی در مدل آرنون | ۱۵ |
| ۳-۳ | نمونه‌های ترمیم تصویر با شبکه‌ی آرنون | ۱۷ |
| ۱-۴ | چهار گونه از روش‌های ترمیم تصویر | ۲۰ |
| ۲-۴ | معماری شبکه‌ی دومسیره | ۲۱ |
| ۳-۴ | روش ترکیب ویژگی‌های میانی مسیر معکوس در مسیر روبه جلو | ۲۲ |
| ۴-۴ | نتایج ترمیم تصویر روش دومسیره روی مجموعه داده‌ی FFHQ | ۲۳ |
| ۱-۵ | مقایسه‌ی نتایج ترمیم تصویر برای سه شبکه‌ی یو با ناحیه‌ی دریافت‌های مختلف | ۲۶ |
| ۲-۵ | معماری کلی شبکه با بهبود محلی و سراسری | ۲۸ |
| ۳-۵ | نتایج ترمیم تصویر شبکه‌ی بهبود محلی و سراسری | ۳۰ |
| ۱-۶ | معماری شبکه‌ی بزرگنمایی و ترمیم | ۳۳ |
| ۲-۶ | ساختار کانولوشن دروازه‌ای | ۳۴ |
| ۳-۶ | نتایج ترمیم تصویر با شبکه‌ی بزرگنمایی و ترمیم | ۳۵ |

فهرست اختصارات

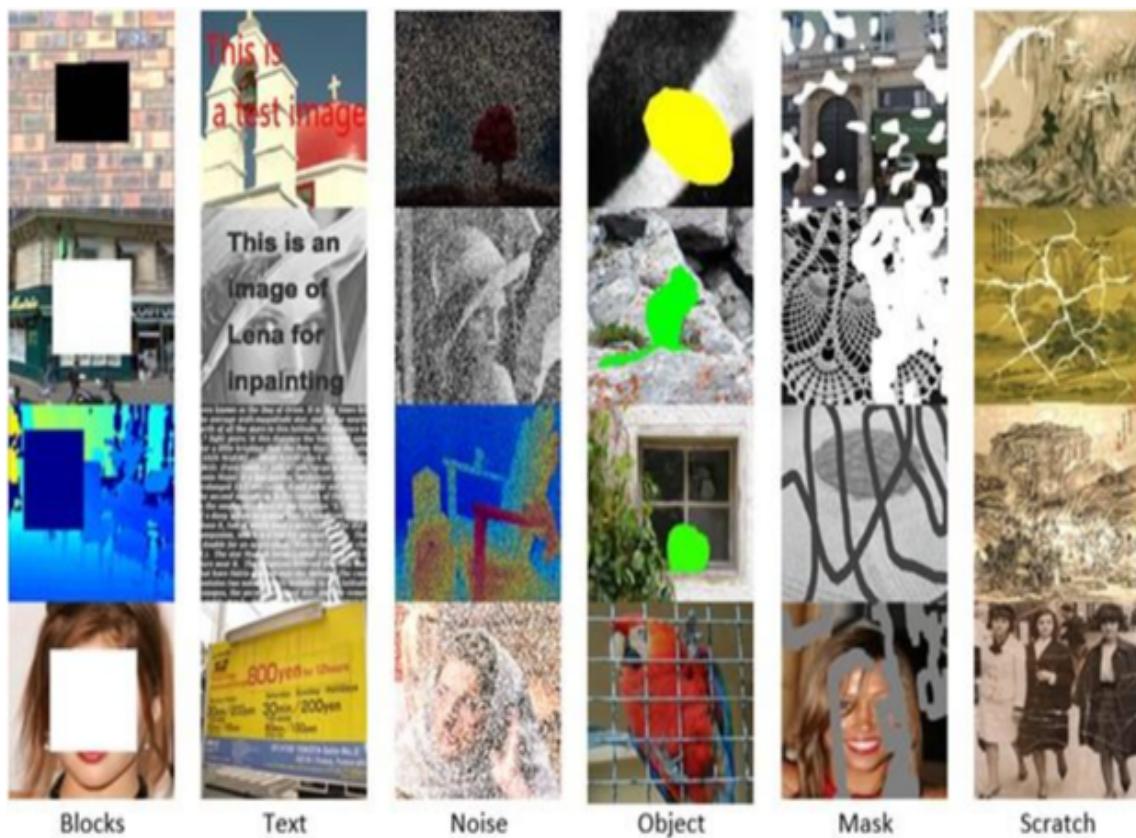
عنوان اختصاری عنوان كامل

تمت تبديل محتوایی تجمیعی

فصل اول

مقدمه

ترمیم تصویر^۱ سال‌های طولانی از مسائل مهم در حوزه‌ی بینایی ماشین بوده است و در سال‌های گذشته با توجه به رشد تکنیک‌های بینایی ماشین اهمیت بیشتری یافته است. ترمیم تصویر عبارتست از پر کردن نواحی از دست رفته یا ماسک‌شده‌ی تصویر به گونه‌ای که طبیعی و مناسب با تصویر باشد. بنابراین ترمیم تصویر باید به گونه‌ای باشد که اولاً با اطلاعات معنایی^۲ مطابقت داشته باشد؛ ثانیاً از نظر بافت، رنگ و جزئیات با تصویر مناسب و هماهنگ باشد. ترمیم تصویر برای انواعی از تخریب‌های متنوع چون متن روی تصویر، نویز، ماسک، خش خوردگی، شی‌های نامربوط و غیره به کار می‌رود. شکل ۱-۱ برخی از تخریب‌های ممکن را نشان می‌دهد. همین گسترده‌گی طیف اصلاحات انجام شده در ترمیم تصویر سبب می‌شود تا در وظایف مختلفی چون بازیابی تصویر^۳، اصلاح تصویر^۴ و کدگذاری و انتقال تصویر به کار گرفته شود.



شکل ۱-۱: انواعی از تخریب‌ها در مساله‌ی ترمیم تصویر

انواعی از تخریب‌ها در تصاویر با توجه به کاربرد ممکن است ایجاد شود که روش‌های ترمیم تصویر قادر به اصلاح آنها و بازسازی تصویر هستند[۳].

¹Image inpainting

²Semantic information

³Image restoration

⁴Image editing

روش‌های موجود در ترمیم تصویر به دو دسته‌ی روش‌های سنتی بدون یادگیری و روش‌های مبتنی بر یادگیری تقسیم می‌شوند. با توجه به سابقه‌ی طولانی مدت وظیفه‌ی ترمیم تصویر در حوزه‌ی بینایی ماشین روش‌های سنتی آن عمدتاً مبتنی بر روش‌های ریاضیاتی و فیزیکی و بر پایه‌ی ویژگی‌های آماری تصویر هستند. این روش‌های سنتی را می‌توان به دو دسته‌ی روش‌های مبتنی بر انتشار^۴ و روش‌های مبتنی بر وصله^۵ تقسیم نمود.

روش‌های مبتنی بر انتشار، ناحیه‌ی از دست رفته را با شروع از یک مرز و انتشار هموار محتویات تصویر بر مبنای یک معادله دیفرانسیل جزئی بازسازی می‌کنند. این رویکردها از معادلات دیفرانسیل مختلفی و یا ویژگی‌های مختلفی چون واریانس بهره می‌گیرند اما همگی آنها تنها در تکمیل نواحی باریک و کوچک از دست رفته و نواحی فاقد اطلاعات معنایی مناسب هستند و در نواحی بزرگ که با دور شدن از مرز ناحیه‌ی از دست رفته اطلاعات محتوایی کاهش می‌یابد عملکرد مناسبی ندارند.

روش‌های مبتنی بر وصله نیز ناحیه‌ی از دست رفته را با جستجو برای بهترین وصله‌ی ممکن در نواحی غیرآسیب دیده‌ی تصویر و کپی کردن آن در ناحیه‌ی از دست رفته، ترمیم می‌کنند. یک روش رایج این دسته، روش تطبیق وصله^۶ است که در آن برای کاهش هزینه‌های تطبیق وصله از یک الگوریتم نزدیکترین همسایه‌ی تصادفی استفاده می‌شود که در کاربردهای اصلاح تصویر بسیار به کار گرفته می‌شوند [۱]. در تصاویری که نواحی پس زمینه‌ای از دست رفته دارند و یا تصاویری که بافت و الگوهای تکراری دارند، این روش عملکرد خوبی دارد اما با توجه به اینکه از اطلاعات خود تصویر برای ترمیم آن استفاده می‌کند، در صورتی که ناحیه‌ی از دست رفته شامل اشکال پیچیده یا الگویی از قبل مشاهده نشده باشد عملکرد این روش‌ها چندان قابل توجه نیست.

با گسترش و توسعه‌ی شبکه‌های عمیق به ویژه شبکه‌های مختص تصویر مانند شبکه‌های کانولوشنی^۷ روش‌های مبتنی بر یادگیری جایگزین روش‌های سنتی شدند. مزیت اصلی این روش‌ها این است که توانایی درک اطلاعات معنایی و محتوایی و استخراج ویژگی‌های سطح بالاتر از تصویر را نیز دارا هستند. روش‌های ترمیم تصویر مبتنی بر یادگیری عمیق را می‌توان به سه دسته‌ی تک مرحله‌ای^۸، دو مرحله‌ای^۹ و روش‌های پیش‌رونده^{۱۱} تقسیم کرد.

مدل‌های اولیه‌ی روش‌های یک مرحله‌ای متشکل از ساختار کدگذار-کدگشا^{۱۲} و توابع هزینه‌ی بازسازی^{۱۳} و مقابله^{۱۴} می‌باشد. از جمله رایج ترین این مدل‌ها، که در زمان ارائه اش با توجه به عملکرد عالی مورد

⁵diffusion-based models

⁶Patch-based methods

⁷PatchMatch

⁸Convolutional neural networks

⁹One-stage

¹⁰Two-stage

¹¹Progressive methods

¹²encoder-decoder

¹³reconstruction loss

¹⁴adversarial loss

توجه قرار گرفت، روش کدگذار محتوایی^{۱۵} است [۵]. این شبکه ترکیبی از شبکه‌های کانولوشنی با یک لایه‌ی تماماً متصل برای کدگذاری محتوایی تصویر در فضای پنهان است. این مدل و تمام مدل‌هایی که بر مبنای آن ساخته شده اند، از دو مشکل رنج می‌برند: اول، عدم توانایی پردازش تصویر با رزولوشن بالا و دوم، عدم توانایی پردازش نواحی از دست رفته با شکل نامشخص به دلیل به کارگیری لایه‌ی تماماً متصل.

برای حل مشکلات موجود در تصاویر با رزولوشن بالا و نیز برای دخیل کردن اطلاعات محتوایی کل تصویر از ایده‌هایی چون ترکیب شبکه‌های تمام کانولوشنی^{۱۶} و مکانیزم توجه^{۱۷} کمک گرفته می‌شود. این ساختارها با توجه به اینکه توانایی دریافت محتوا از نواحی دور از ناحیه‌ی از دست رفته و مدل کردن اطلاعات غیر محلی را دارند عملکرد بهتری از حیث ترمیم معنایی دارند اما دارای ساختارهای تخریب شده و الگوهای تکراری در ناحیه‌ی از دست رفته هستند. روش دیگری که برای به کارگیری اطلاعات معنایی نواحی دور از ناحیه‌ی از دست رفته ارائه شد، استفاده از کانولوشن‌های گسترش یافته^{۱۸} بود که از هسته‌های کانولوشنی که گسترش داده شده و دارای ناحیه‌ی دریافت^{۱۹} بیشتری است استفاده می‌کند [۶]. این شبکه توانایی مشاهده‌ی نواحی بیشتری از تصویر را دارد و درنتیجه استنتاج محتوایی بهتری انجام می‌دهد اما مشکل اساسی آن این است که قادر به دریافت الگوهای از پیش تعریف شده است و قابلیت تشخیص الگوهای مناسب برای استنتاج محتوایی را ندارد.

روش‌های دو مرحله‌ای از یک چارچوب دو مرحله‌ای برای ترمیم تصویر استفاده می‌کند. در این چارچوب یک شبکه برای ساخت کلیات درشت^{۲۰} تصویر و یک شبکه برای بهبود^{۲۱} به کار می‌رود. در این مدل‌ها شبکه‌ی اول اطلاعاتی مانند نقشه‌ی لبه، نقشه‌ی الگوی باینری و کلیات تصویر در قسمت از دست رفته را بازسازی می‌کند و سپس شبکه‌ی بهبود این اطلاعات اولیه و ویژگی‌های درشت مقیاس را بهبود می‌دهد و با اطلاعات محتوایی ترکیب می‌کند.

در نهایت، در روشهای پیش‌رونده از یک روند تکراری برای ترمیم تصویر استفاده می‌شود. به این صورت که در هر گام تنها یک ردیف از پیکسل‌های ناحیه‌ی از دست رفته بازیابی می‌شوند. این روشهای با توجه به ماهیت تکراری آنها هزینه محاسباتی زیادی دارند و علاوه براین در بسیاری از وظایفی که نیاز به ترمیم نواحی نامنظم است، این شبکه‌ها ناموفقند.

در این گزارش به پنج روش متاخر مطرح شده در زمینه‌ی ترمیم تصویر می‌پردازیم. هر کدام از این روشهای به مطرح کردن چالشی حل نشده در بحث ترمیم تصویر ویا ارائه‌ی رویکردي نوین برای بهبود کیفیت خروجی چه از لحاظ ویژگی‌های ظاهری و چه از لحاظ ویژگی‌های معنایی می‌پردازند.

¹⁵Context encoder

¹⁶fully convolutional networks

¹⁷Attention mechanism

¹⁸dilated convolution

¹⁹Receptive field

²⁰Coarse

²¹Refinement

فصل دوم

تبدیل‌های محتوایی تجمعی در شبکه‌های مولد قابلی

۱-۲ ایده‌ی اصلی

بسیاری از روش‌های ترمیم تصویر از شبکه‌های مولد تقابلی برای ترمیم ناحیه‌ی از دست رفته استفاده می‌کنند؛ این روش‌ها علیرغم پیشرفت قابل توجهشان از مشکلاتی چون ساختارهای تخریب شده و بافت‌های محو شده در تصاویر با رزولوشن بالا رنج می‌برند. در واقع با افزایش رزولوشن اکثر روش‌های تاکنون ارائه شده می‌توانند در ارائه‌ی جزئیات با مشکل رو برو شوند. این چالش‌ها همانطور که در فصل قبل هم اشاره شد از دو منشا اصلی سرچشم می‌گیرند: یک، استنتاج محتوایی تصویر با توجه به نواحی دور از ناحیه‌ی از دست رفته، و دوم، تولید بافت مناسب برای نواحی از دست رفته بزرگ.

همانطور که در فصل قبل نیز بحث شد، به کمک روش‌های مبتنی بر مکانیزم توجه و یا کانولوشن‌های گسترش یافته می‌توان اطلاعات دور از ناحیه‌ی از دست رفته را دخیل کرد تا تجمیع این اطلاعات معنایی در ساخت ناحیه‌ی از دست رفته مفید واقع شود. اما هر کدام از روش‌های پیشین محدودیت‌هایی چون ایجاد الگوهای تکراری و استفاده از الگوهای مشخص دارند. روش ارائه شده در این فصل با به کارگیری بلوک تبدیل محتوایی تجمیعی^۱ با به اختصار تمت در صدد حل این چالش است.

بلوک‌های تمت از استراتژی تقسیم-تبدیل-ادغام^۲ استفاده می‌کنند به این صورت که ابتدا هسته‌های یک کانولوشن استاندارد را به تعدادی زیرهسته^۳ تقسیم می‌کنند و در هر زیر هسته از گسترش^۴ متفاوتی استفاده می‌کنند و سپس آنها را به ورودی اعمال می‌کنند و در نهایت از ترکیب این تبدیلات انجام شده، مجموعه ویژگی نهایی را می‌سازند. این بلوک‌ها به واسطه‌ی اعمال هسته‌های گسترش یافته با اندازه‌های مختلف اطلاعات معنایی دور و نزدیک را در مجموعه ویژگی لحاظ می‌کنند.

چالش دومی که در روش‌های پیشین وجود دارد مشکل تولید بافت و ساختار مناسب برای نواحی از دست رفته‌ی بزرگ در تصاویر با رزولوشن بالاست. در روش ارائه شده در این فصل برای تولید بافت‌ها و ساختارهای مناسب شبکه‌ی تمايزگر^۵ شبکه‌ی مولد تقابلی اصلاح شده است. در شبکه‌های مولد تقابلی پیشین که به منظور ترمیم تصویر به کار گرفته می‌شدند، از تمايزگر از پیش آموزش داده شده شبکه مولد تقابلی وصله‌ای^۶ برای تمايزگر استفاده می‌شود. این شبکه برای تک تک وصله‌های تصویر عمل تمايز واقعی از جعلی را انجام می‌دهد و نقطه ضعف این روش نیز همین موضوع است چرا که در ترمیم تصویر با تولید کامل یک تصویر طرف نیستیم بلکه با بخش‌هایی از تصویر رو برو هستیم که متعلق به تصویر ترمیم شده است و بنابراین ستuzzi روی آن انجام نشده که نیازمند بازخورد تمايزگر باشد. در مدل تمت در شبکه‌های مولد تقابلی از یک تمايزگر پیش بینی ماسک^۷ استفاده شده است که تمرکز آن روی وصله‌های سنتز شده است و بازخورد آن به شبکه کمک کننده کمک بیشتری می‌کند.

¹ Aggregated contextual transformation

² split-transform-merge

³ Sub-kernel

⁴ dilation

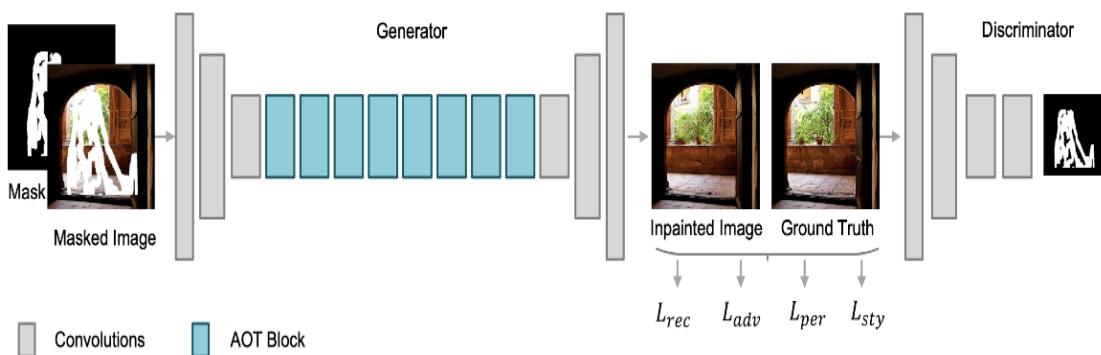
⁵ Discriminator

⁶ PatchGAN

⁷ mask-prediction discriminator

۲-۲ معماری

روش تبدیل محتوایی تجمیعی با شبکه‌ی مولد تقابلی^۱ با هدف تولید محتوای منطقی در کنار بافت واضح برای تصاویر با رزولوشن بالا ارائه شد. معماری این شبکه در شکل اولیه‌ی خود معماری یک شبکه‌ی مولد تقابلی است که از دو شبکه‌ی تولیدکننده و تمایزگر ایجاد شده است. شکل کلی معماری این شبکه در تصویر ۱-۲ آمده است. شبکه‌ی تولیدکننده شامل سه بخش کدگذاری تصویر ورودی ساخته می‌شود و کدگذار از روی هم گذاشتن تعدادی لایه‌ی کانولوشن برای کدگذاری تصویر ورودی ساخته می‌شود و سپس این تصویر کدگذاری شده به بلوک‌های متوالی تمت داده می‌شود و سپس از یک کدگشا که از روی هم گذاشتن تعدادی کانولوشن معکوس ساخته می‌شود عبور می‌کند. بنابراین مهم ترین قسمت شبکه‌ی تولیدکننده بلوک تمت می‌باشد.



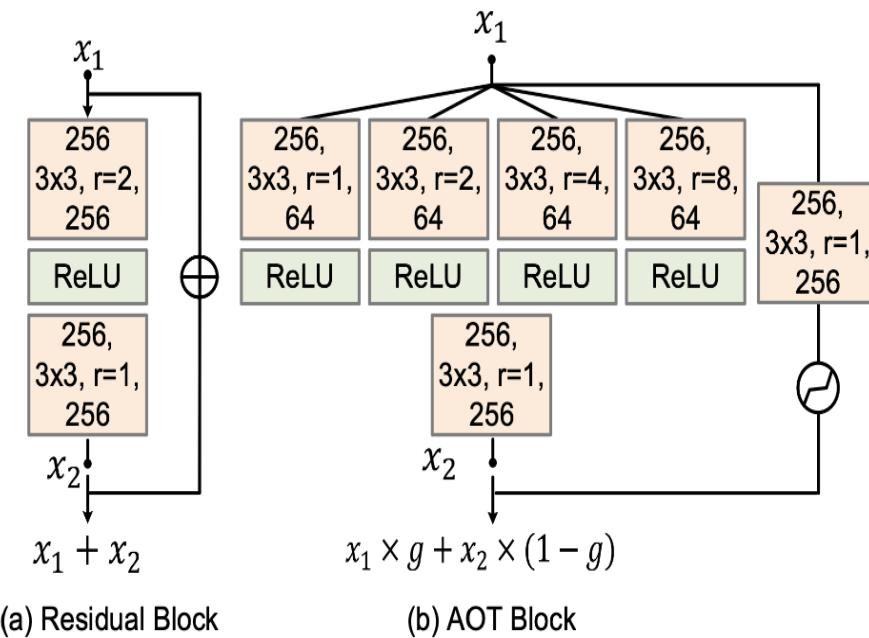
شکل ۱-۲: معماری تمت با شبکه‌ی مولد تقابلی

این شبکه مانند سایر شبکه‌های مولد تقابلی از دو زیرشبکه‌ی تولیدکننده و تمایزگر تشکیل شده است [۱۰].

بلوک تمت که در شکل ۲-۲ نشان داده شده است، از استراتژی تقسیم-تبدیل-ادغام استفاده می‌کند. ابتدا هسته‌های یک کانولوشن استاندارد را به تعدادی زیر هسته هر کدام با تعداد کanal خروجی کمتر تقسیم می‌کند. به عنوان مثال یک هسته با ۲۵۶ کanal خروجی را به ۴ هسته با ۶۴ خروجی تقسیم می‌کند. سپس روی هر مجموعه از هسته‌ها گسترش با نرخ متفاوتی انجام می‌دهد. عمل گسترش در هسته به معنای قرار دادن صفر بین دو خانه‌ی مجاور آن است و هر چقدر که نرخ گسترش بیشتر باشد، زیرهسته‌ی مربوطه قسمت‌های بیشتری از تصویر را می‌بینند. در بخش ادغام بلوک تمت نیز ویژگی‌های به دست آمده با زیرهسته‌های مختلف به یکدیگر الحق می‌شوند و سپس یک کانولوشن استاندارد روی آنها اعمال می‌شود. هدف اصلی بلوک تمت این است که به شبکه‌ی تولیدکننده اجازه دهد تا الگوهای متعدد را از مسیرهای مختلف استخراج و دریافت کند و از آنها برای استخراج معنایی استفاده کند.

معماری شبکه‌ی تمایزگر همان معماری شبکه‌ی مولد تقابلی وصله‌ای است که از تعدادی لایه‌ی کانولوشن استاندارد که هر کدام اندازه‌ی نقشه‌ی ویژگی را نصف می‌کنند ساخته شده است. این شبکه

⁸AOT-GAN



شکل ۲-۲: ساختار بلوک تبدیل محتوای تجمیعی

شکل سمت چپ یک بلوک باقیمانده‌ای ساده را نشان می‌دهد که به طور معمول در شبکه‌های مولد تقابلی به کار برده می‌شود و شکل سمت راست یک بلوک تمت را نشان می‌دهد [۱۰].

تصویر تولید شده توسط تولیدکننده و یا داده‌ی حقیقی را می‌گیرد و یک نقشه‌ی پیش‌بینی بر می‌گرداند که شامل پیش‌بینی حقیقی یا جعلی بودن هر وصله توسط تمایزگر است. شکل ۱-۲ این ساختار را نشان می‌دهد.

همانطور که پیش‌تر هم اشاره شد، مشکل اصلی این تمایزگر این است که برای نواحی آسیب ندیده‌ی تصویر نیز پیش‌بینی انجام می‌دهد و این پیش‌بینی‌ها را به صورت بازخورد به شبکه‌ی تولید کننده می‌دهد. برای حل این مشکل در هر مرحله تنها وصله‌هایی که توسط شبکه‌ی تولید کننده تولید شده‌اند و پیش‌بینی آنها به صورت بازخورد به تولیدکننده داده می‌شوند و به این منظور از یک ماسک برای تمیز نقاط از دست رفته از نقاط سالم تصویر استفاده می‌شود تا فقط بازخوردهای نواحی آسیب‌دیده به تولیدکننده برسد.

۳-۲ آموزش شبکه

در این شبکه همانند اکثر شبکه‌های ترمیم تصویر از چهار تابع هزینه‌ی بازسازی^۹، تقابلی^{۱۰}، استایل^{۱۱} و ادراکی^{۱۲} استفاده شده است. خطای بازسازی میزان تابع هزینه‌ی L_1 را در تفاضل پیکسل به پیکسل عکس واقعی و عکس ترمیم شده محاسبه می‌کند. خطای ادراکی و خطای استایل هر دو از خطاهای بسیار پرکاربرد در ترمیم تصویر هستند. خطای ادراکی فاصله‌ی L_1 بین نقشه‌ی فعالسازی یک شبکه‌ی پیش آموزش داده شده روی عکس‌های حقیقی و ترمیم شده را محاسبه می‌کند که در این مقاله شبکه‌ی مورد استفاده، شبکه‌ی $VGG-19$ ^{۱۳} است. به طور مشابه خطای استایل نیز فاصله‌ی L_1 بین ماتریس گرام ویژگی‌های عمیق تصاویر ترمیم شده و اصلی را محاسبه می‌کند. خطای ادراکی تضمین می‌کند که دو تصویر از نظر محتوایی و ویژگی‌های استخراجی مشابه باشند و خطای استایل نیز تضمین می‌کند که دو تصویر استایل یکسانی داشته باشند. در نهایت خطای تقابلی نیز برای آموزش شبکه‌های مولد تقابلی به کار می‌رود و شامل بهینه کردن هر دو شبکه‌ی تولیدکننده و تمایزگر می‌باشد. تابع هدف نهایی شبکه به صورت زیر می‌باشد:

$$L = \lambda_{adv} L_{adv}^G + \lambda_{rec} L_{rec} + \lambda_{per} L_{per} + \lambda_{sty} L_{sty} \quad (1-2)$$

۴-۲ نتایج

تصویر ۳-۲ نتایج به دست آمده در ترمیم تصویر برای شبکه‌ی مطرح شده و مجموعه‌ای از بهترین روش‌های ترمیم تصویر را نشان می‌دهد. این تصویر پیش از همه نشانگر این موضوع است که در تصاویر با رزولوشن بالا روش تمت با شبکه‌های مولد تقابلی قابلیت بازسازی جزئیات مناسب با معنا و دارای بافت متناسب را دارد. همچنین نتایج ارائه شده در مقاله نشان می‌دهد که در درصدهای مختلفی از تخریب و با معیارهای مختلف سنجش کیفیت و دقت بازسازی روش ارائه شده نسبت به سایر روش‌های پیشین برتری دارد.

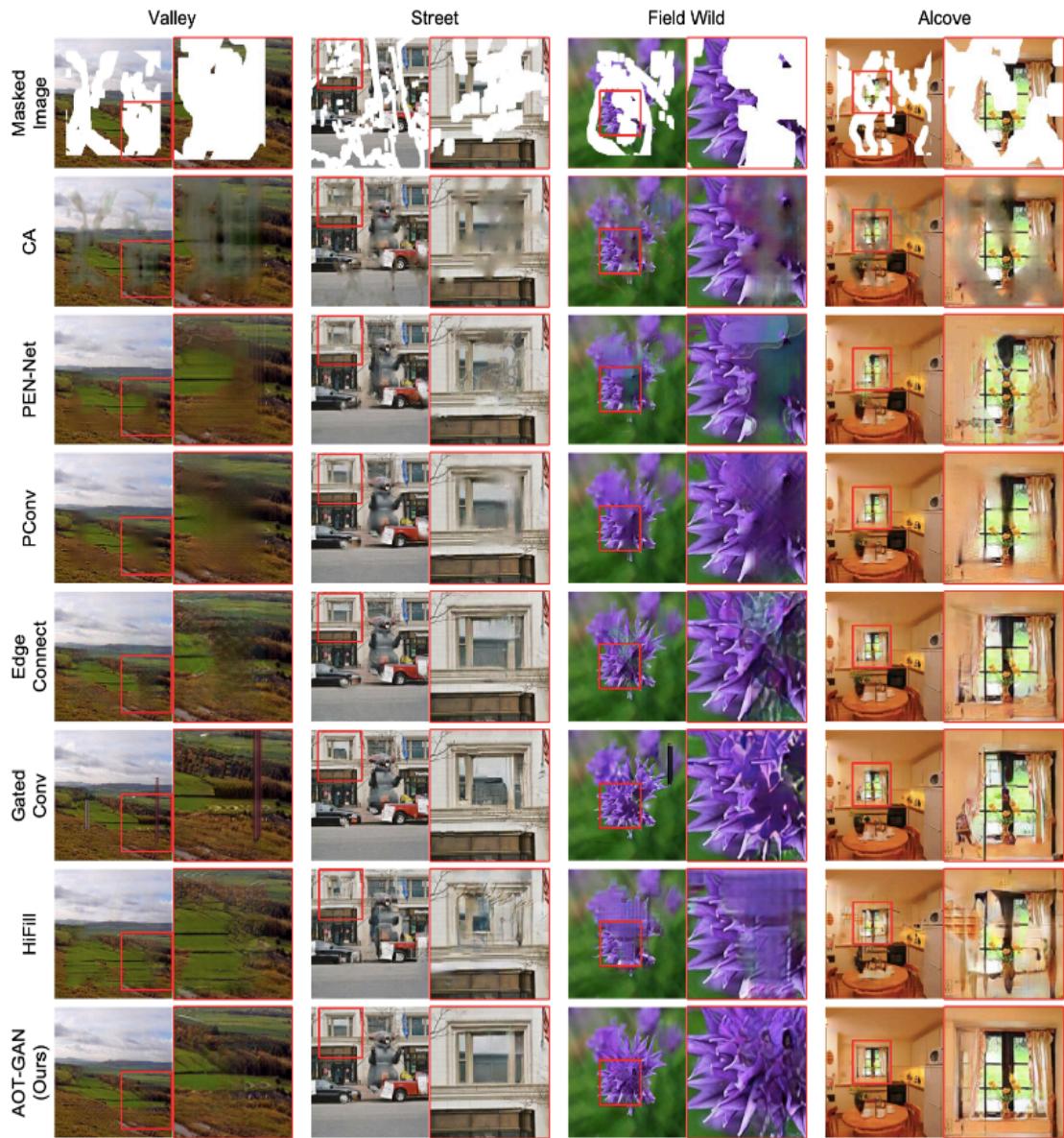
⁹Reconstruction loss

¹⁰Adversarial loss

¹¹Style loss

¹²Perceptual loss

¹³Gram matrix



شکل ۲-۳: نتایج ترمیم تصویر در روش تمت با شبکه‌ی مولد تقابلی و سایر شبکه‌ها
این نتایج به خوبی نشان می‌دهند که در ایجاد بافت‌های با جزئیات روش تمت با شبکه‌ی مولد تقابلی نسبت به شبکه‌های پیشین خود عملکرد قابل توجهی دارد [۱۰].

فصل سوم

شبکه‌ی تعمیر و شبکه‌ی بهینه سازی

۱-۳ ایده‌ی اصلی

در بسیاری از روش‌های یک مرحله‌ای مبتنی بر شبکه‌های کانولوشنی با پدیده‌ای به نام انحراف محلی رنگ^۱ مواجهیم. این پدیده عبارتست از عدم هماهنگی در رنگ یا طیف تغییرات رنگی ناحیه‌ی از دست رفته با نواحی مجاور. علاوه بر این پدیده، مرزهای مشخص و بافت‌های فازی از دیگر مشکلاتی هستند که شبکه‌ی تعمیر و شبکه‌ی بهینه سازی که به آن شبکه‌ی آرنون نیز می‌گوییم در صدد برطرف کردن آن برآمده‌اند. نویسنده‌گان مقاله ادعا می‌کنند که روش‌های پیشین مبتنی بر محتوا یا صرفاً شبکه‌های کانولوشنی به خوبی اطلاعات محتوا ای را دخیل می‌کنند و تصویر را ترمیم می‌کنند اما در تصویر ترمیم شده همچنان ناهمانگی در رنگ و بافت وجود دارد. شبکه‌ی ارائه شده از یک رویکرد دو مرحله‌ای استفاده می‌کند و در مرحله‌ی اول با استفاده از یک شبکه‌ی تعمیر تلاش می‌کند تا رنگ و مکان هر پیکسل تا حد امکان به رنگ و مکان پیکسل در تصویر واقعی نزدیک باشد و بعد در مرحله‌ی دوم با استفاده از شبکه‌ی بهینه سازی ناپیوستگی‌ها در رنگ و بافت را اصلاح و برطرف می‌کنند.

۲-۳ معماری

۱-۲-۳ شبکه‌ی تعمیر

همانطور که در بخش ایده هم بیان شد وظیفه‌ی اصلی شبکه‌ی تعمیر ارائه‌ی یک مجموعه پیکسل متناسب با آنچه در تصویر اصلی وجود دارد می‌باشد. به عبارتی این شبکه با به کارگیری اطلاعات محتوا ای در کنار اطلاعات محلی تصویر سعی در بازسازی ناحیه‌ی از دست رفته دارد.

در معماری این شبکه از یک شبکه‌ی مولد تقابلی شامل شبکه‌ی تولیدکننده و شبکه‌ی تمایزگر استفاده شده است. شبکه‌ی تولیدکننده از ساختار کانولوشن جزئی^۲ در کنار یک شبکه‌ی یو^۳ استفاده شده است. شکل ۱-۳ قسمت a معماری شبکه‌ی تولیدکننده را نشان می‌دهد.

کانولوشن جزئی عملیاتی است که در وظیفه‌ی ترمیم تصویر بسیار به کار گرفته می‌شود. برخلاف کانولوشن استاندارد که تمامی پیکسل‌های داخل پنجره‌ی هسته را در نظر می‌گیرد، در کانولوشن جزئی تنها پیکسل‌هایی که دارای مقدار معتبر هستند در نظر گرفته می‌شوند؛ یعنی پیکسل‌های خارج از مرز و پیکسل‌هایی از تصویر که به ناحیه‌ی از دست رفته تعلق دارند را در محاسبات در نظر نمی‌گیرد. استفاده از کانولوشن‌های جزئی به ترمیم تصویر و یادگرفتن جزئیات آن کمک می‌کند و لایه‌ی نرمال‌سازی دسته‌ای^۴ نیز در افزایش سرعت یادگیری شبکه و توانایی تعمیم آن موثر است.

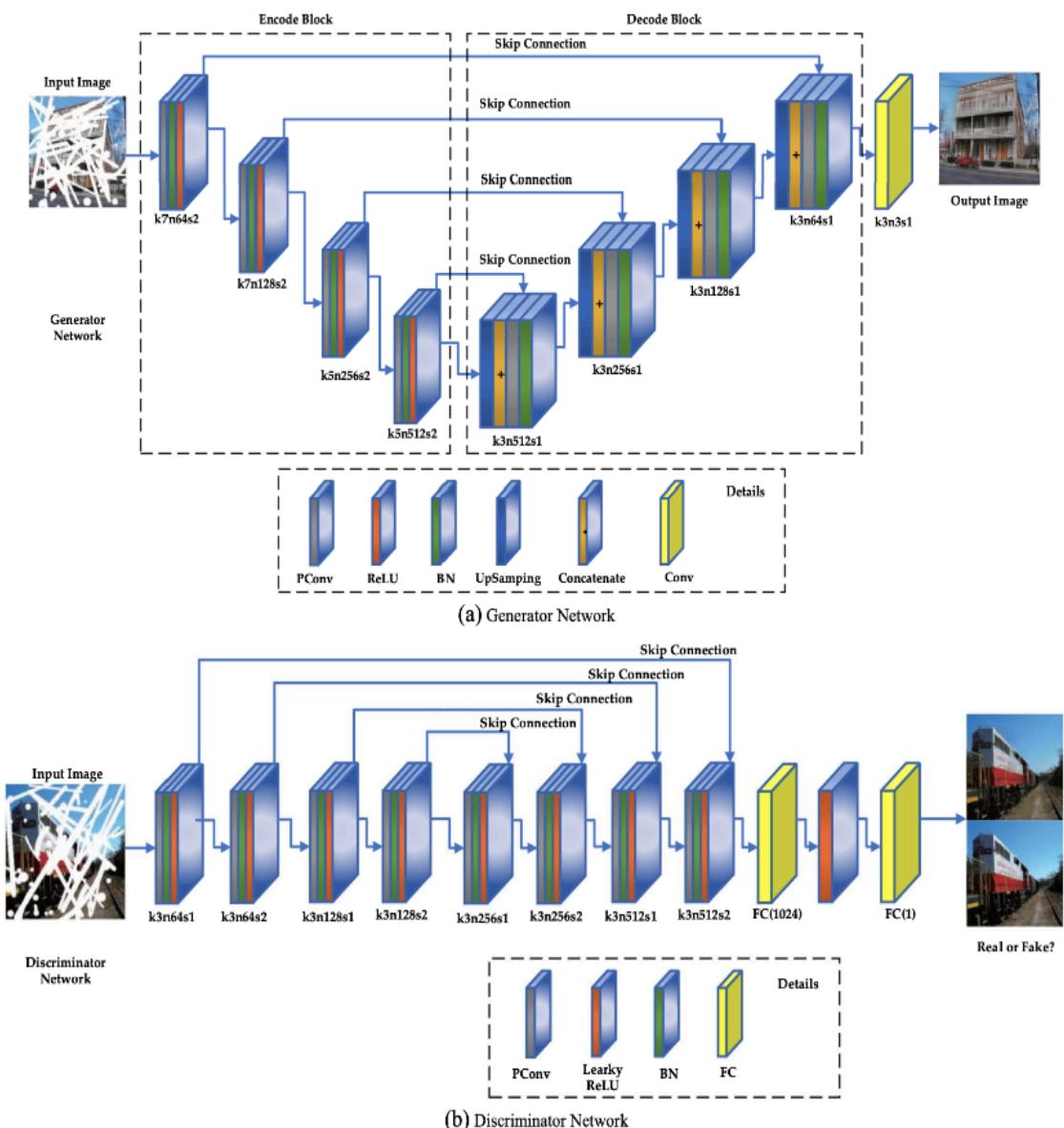
شبکه‌ی تولیدکننده ساختاری مشابه شبکه‌ی یو دارد که از یک کدگذار و یک کدگشا تشکیل شده

¹Local chromatic aberration

²partial convolution

³U-net

⁴Batch normalization



شکل ۳-۱: معماری شبکه‌ی تعمیر در مدل آرنون

معماری شبکه‌ی تعمیر که در فرم کلی یک شبکه‌ی مولد تقابلی است و شامل یک شبکه‌ی تولیدکننده و یک شبکه‌ی تمایزگر است. قسمت a معماری شبکه‌ی تولیدکننده و قسمت b معماری شبکه‌ی تمایزگر را نشان می‌دهد [۲].

است. در کدگذار از کانولوشن جزوی استفاده شده است و در کدگشا از بالانمونه‌گیری^۵ استفاده شده است. همچنین برای تجمعی اطلاعات معنایی سطح بالا و اطلاعات مکانی و محلی سطح پایین از اتصالات پرشی میان لایه‌های متناظر کدگذار و کدگشا استفاده شده است. در نهایت نیز برای ارائه‌ی خروجی از یک کانولوشن با اندازه‌ی هسته‌ی یک و گام یک برای کاهش تعداد کانال‌ها و نیز یک تابع فعلسازی

⁵Upsampling

سیگموید^۶ استفاده شده است.

شبکه‌ی تمایزگر نیز شامل هشت بلوک کانولوشن و دولایه‌ی تماماً متصل است که ساختار آن در شکل ۱-۳ قسمت b نشان داده شده است.

۲-۲-۳ شبکه‌ی بهینه سازی

با توجه به اینکه عکس بازسازی شده توسط شبکه‌ی تعمیر حاوی درجه‌ای از انحراف محلی رنگ است، وظیفه‌ی اصلی شبکه‌ی بهینه سازی هموارسازی رنگ و ساختار بازسازی شده و حذف مصنوعات در اطراف مرزها و ناسازگاری‌ها و ناهماهنگی هاست. این شبکه نیز مشابه شبکه‌ی تعمیر دارای ساختار شبکه‌های مولد تقابلی است. شبکه‌ی تولیدکننده از تعدادی بلوک چند مقیاسی باقی‌مانده‌ای^۷ تشکیل شده است که هر کدام از این بلوک‌ها نیز از دو قسمت تشکیل شده‌اند. قسمت اول شامل ۴ بلوک کانولوشن گسترش داده شده است که نرخ گسترش آنها به ترتیب ۱، ۲، ۴ و ۸ است. استفاده از کانولوشن‌های گسترش داده شده قابلیت استخراج ویژگی از ناحیه‌های دریافت مختلف را فراهم می‌آورد که باعث یادگیری معنایی چند مقیاسی و افزایش توانایی استخراج ویژگی می‌شود. بخش دوم این بلوک‌ها نیز بلوک‌های اتصالات باقی‌مانده‌ای هستند که کارکرد آنها الصاق ویژگی‌های استخراج شده از کانولوشن‌های گسترش یافته و ترکیب آنها به واسطه‌ی یک کانولوشن دیگر است. ساختار این شبکه در شکل ۲-۳ قسمت a آمده است. شبکه‌ی تمایزگر شبکه‌ی بهینه‌سازی نیز ساختاری کاملاً مشابه شبکه‌ی تمایزگر شبکه‌ی تعمیر دارد. تنها تفاوت آن اضافه شدن دو بلوک کانولوشن است که به تشخیص بهتر تغییرات کوچک کمک می‌کند. ساختار این تمایزگر در شکل ۲-۳ قسمت b آورده شده است.

۳-۳ آموزش شبکه

۱-۳-۳ توابع خطای شبکه‌ی تعمیر

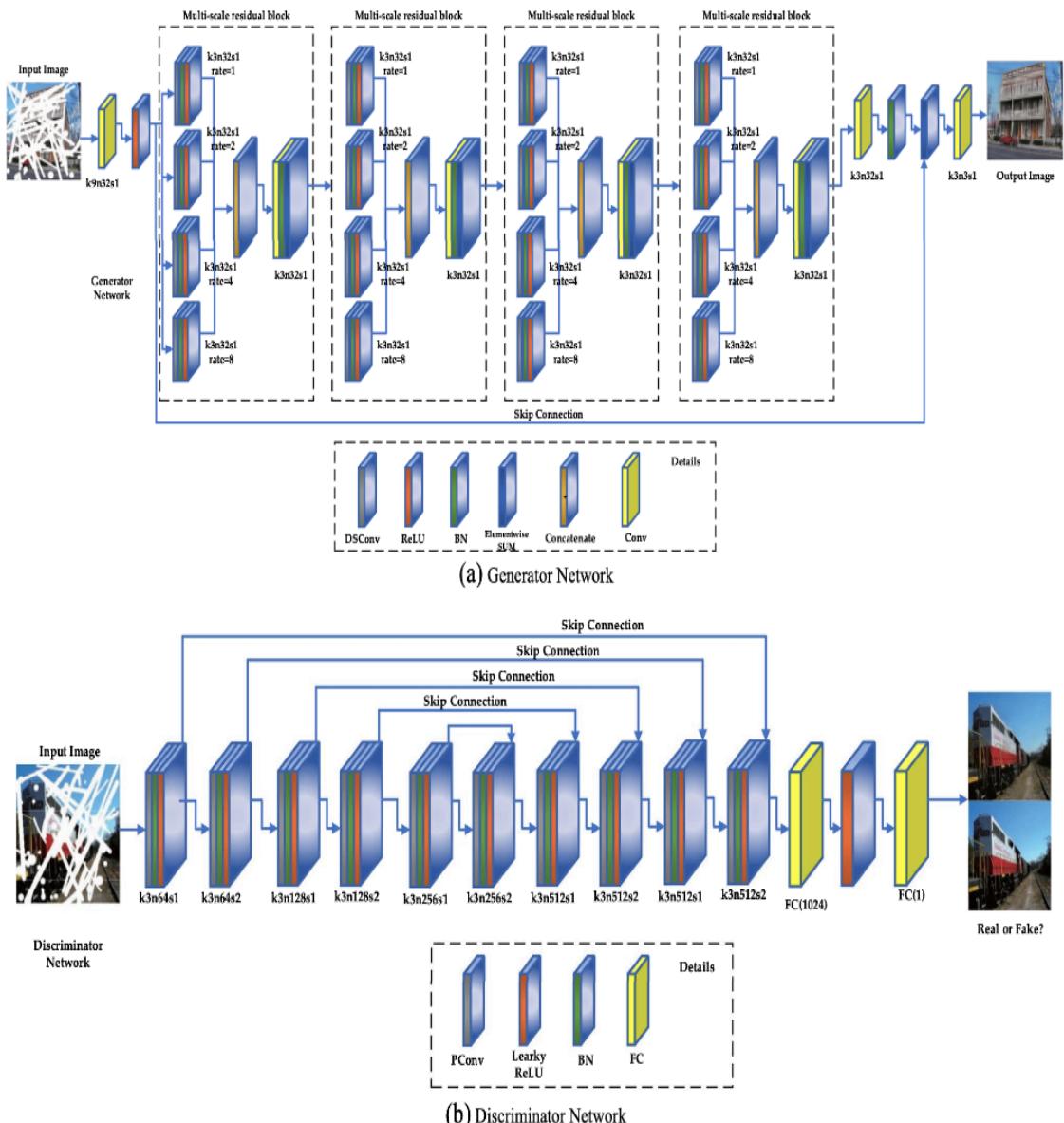
خطای کلی شبکه‌ی تعمیر در رابطه‌ی (۱-۳) آورده شده است. این خطا شامل خطای بازسازی قسمت ماسک شده، خطای بازسازی قسمت ماسک نشده، خطای ادراکی، خطای استایل، خطای تقابلی و خطای تغییرات کلی^۸ می‌باشد.

$$L_{total}^{inp} = 2L_{valid} + 12L_{hole} + 0.04L_{per} + 100(L_{style}^1 + L_{style}^2) + 100L_{adv} + 0.3L_{var} \quad (1-3)$$

⁶Sigmoid

⁷Multi-scale residual block

⁸Total variation



شکل ۳-۲: معماری شبکه‌ی بهینه‌سازی در مدل آرنون

معماری شبکه‌ی بهینه‌سازی که در فرم کلی یک شبکه‌ی مولد تقابلی است و شامل یک شبکه‌ی تولیدکننده و یک شبکه‌ی تمایزگر است. قسمت *a* معماری شبکه‌ی تولیدکننده و قسمت *b* معماری شبکه‌ی تمایزگر را نشان می‌دهد [۲].

خطای بازسازی قسمت ماسک نشده از خطای L_1 تفاضل قسمت از دست نرفته‌ی تصویر در تصویر ترمیم شده و تصویر اصلی به دست می‌آید. این خطا اطمینان حاصل می‌کند که بخش‌های ماسک نشده تغییری نکنند. خطای بازسازی ماسک شده نیز از همین رابطه بر روی قسمت ماسک شده به دست می‌آید که اطمینان حاصل می‌کند که بخش‌های از دست رفته تا حد امکان مشابه تصویر اصلی باشند. خطای ادراکی و خطای استایل همانطور که در فصل قبل اشاره شد با هدف ایجاد هماهنگی در استایل و

ویژگی‌های مفهومی تصویر ترمیم شده با تصویر اصلی محاسبه می‌شوند که در این شبکه برای محاسبه آنها از ویژگی‌های دریافت شده از شبکه‌ی $VGG - 16$ استفاده می‌شود. خطای تقابلی برای آموزش تقابلی شبکه مورد استفاده قرار می‌گیرد و هدف آن این است که هم تولیدکننده و هم تمایزگر خروجی بهینه‌تری تولید کنند فرمول خطای تقابلی برای این شبکه به صورت زیر است:

$$L_{adv} = \frac{1}{N} \sum_{i=0}^{N-1} |D_{inp}(I_{inp}(x_i)) - D_{inp}(I_{real}(x_i))| \quad (2-3)$$

در نهایت خطای تغییرات کلی برای جلوگیری از تغییرات ناگهانی ناحیه‌ی از دست رفته می‌باشد و هموار بودن این ناحیه را تضمین می‌کند.

۲-۳-۳ توابع خطای شبکه‌ی بهینه سازی

هدف توابع این شبکه باید نگهداری نواحی بازسازی شده ای که ساختار مناسبی دارند در کنار بهبود نواحی دارای انحراف محلی رنگ باشد. تابع هدف این شبکه شامل خطای محتوا^۹، خطای مفهومی و خطای تقابلی است که به صورت زیر می‌باشد

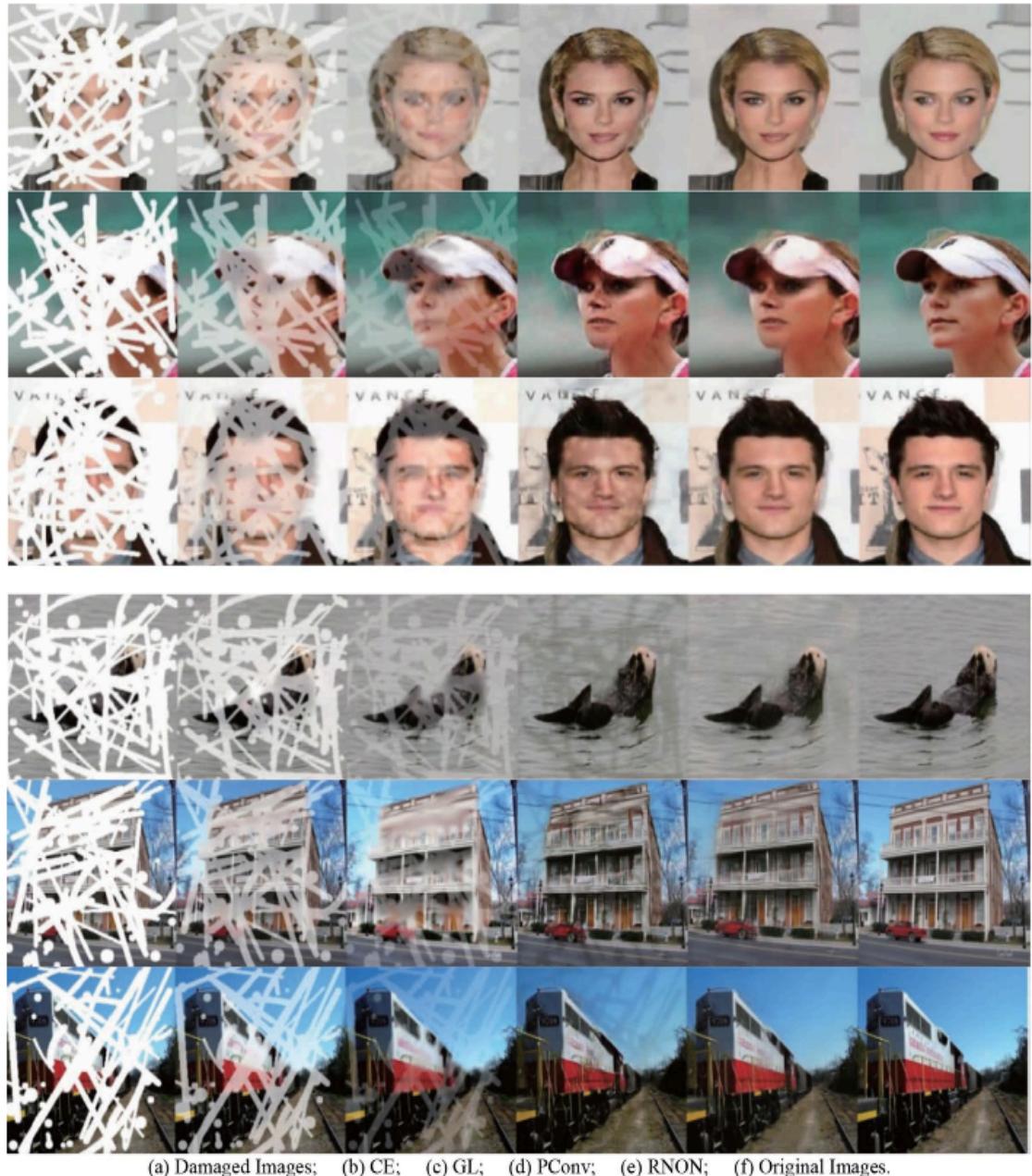
$$L_{total}^{opt} = 40L_{con} + L_{per} + 0.75L_{adv} \quad (3-3)$$

خطای محتوا جمع وزن دار خطای تصویر اصلی و تصویر ترمیم شده است که با ماسک تصویر وزن دهی می‌شود. خطای ادراکی هدفی مشابه قسمت قبل دارد با این تفاوت که در محاسبه‌ی آن از $VGG - 19$ استفاده شده است و خطای تقابلی نیز مشابه شبکه‌ی تعمیر به هدف بهبود هردو شبکه‌ی تولیدکننده و تمایزگر تعریف می‌شود.

۴-۳ نتایج

روش ارائه شده بر روی دو مجموعه داده‌ی *CelebA* و *Places2* آموزش و ارزیابی شده است. و با روش‌هایی چون انکدر محتوایی [۵] و کانولوشن جزئی مقایسه شده است. نتایج این مقایسه بر روی هر دو مجموعه داده در شکل ۳-۳ آورده شده است. همانطور که در شکل هم مشاهده می‌شود، تمامی روش‌ها از نظر معنایی عملکرد خوبی دارند اما از نظر ارائه‌ی جزئیات و حذف ساختارهای تار و محو شبکه‌ی آرنون محسودگی کمتر و جزئیات بیشتری دارد.

⁹content loss



شکل ۳-۳: نمونه‌های ترمیم تصویر با شبکه‌ی آرnon

شکل نمونه‌هایی از ترمیم تصویر با شبکه‌ی آرnon در مقایسه با سایر روش‌های ترمیم تصویر را نشان می‌دهد. همانطور که مشاهده می‌کنید این شبکه در بازسازی جزئیات مانند چشم و یا جزئیات دور مانند پنجره‌ی ساختمان عملکرد مطلوبی دارد [۲].

فصل چهارم

ترمیم تصویر دومسیره

۱-۴ ایده‌ی اصلی

در این فصل تمرکز شبکه‌ی ارائه شده بر روی حل مشکل نواحی از دست رفته‌ی بزرگ است. با وجود عملکرد مناسبی که شبکه‌های کانولوشنی و شبکه‌های مولد تقابلی دارند، در تکمیل نواحی از دست رفته‌ی بزرگ که الگوی معنایی پیچیده‌ای دارند عملکرد مناسبی از خود ارائه نمی‌دهند. یک رویکرد برای حل این مشکل استفاده از معکوس سازی شبکه‌های مولد تقابلی^۱ است. این دسته از روش‌ها معمولاً از یک مدل پیش آموزش دیده‌ی مولد(مانند شبکه‌ی مولد تقابلی استایل^۲) برای تولید دانش پیشین معنایی استفاده می‌کنند. این روش به دنبال کد پنهان^۳ معنایی هستند که نزدیکترین کد به کد تصویر تخریب شده در فضای پنهان باشد و سپس این کد را معکوس می‌کنند تا یک تصویر کامل به کمک شبکه‌ی تولید کننده از پیش آموزش داده شده بسازند.

رویکردهای متفاوتی برای به دست آوردن کد پنهان وجود دارد که می‌توان آنها را به دو دسته‌ی مبتنی بر آموزش و مبتنی بر بهینه سازی تقسیم کرد. در رویکرد مبتنی بر بهینه سازی به صورت گام به گام کد پنهان را به روز رسانی می‌کنیم تا خطای بازسازی با کد پنهان کمینه شود. اما در روش‌های مبتنی بر یادگیری از یک شبکه‌ی کد گذار برای به دست آوردن کد پنهان بر حسب تصویر تخریب شده استفاده می‌کنیم. شکل ۱-۴ این دو روش را نشان می‌دهد. مشکل روش‌های مبتنی بر بهینه سازی هزینه محاسباتی زیاد آن و مشکل روش‌های مبتنی بر یادگیری ارائه‌ی نتایج ضعیف است.

در شبکه‌ی ارائه شده در این فصل از مزایای رویکردهای معکوس سازی شبکه‌های مولد تقابلی در کنار یک شبکه‌ی رو به جلو^۴ برای ترمیم تصویر استفاده می‌شود. در این شبکه دو مسیر وجود دارد یک مسیر روبه جلو^۵ و یک مسیر معکوس سازی^۶. در مسیر معکوس سازی ما به دنبال یک کد پنهان می‌گردیم که تصویر متناظر آن نزدیکترین تصویر به تصویر تخریب شده باشد و با توجه به زمانبند بودن روش‌های مبتنی بر بهینه سازی از روش‌های مبتنی بر یادگیری استفاده می‌کنیم. سپس اطلاعات به دست آمده و ویژگی‌های استخراج شده در مسیر معکوس را در مسیر رو به جلو به کار می‌گیریم. در مسیر روبه جلو از یک شبکه‌ی خودکدگذار^۷ استفاده شده است که ویژگی‌های مقیاس‌های مختلف کدگشای مسیر معکوس را دریافت می‌کند و عکس را ترمیم می‌کند.

¹GAN inversion

²StyleGAN

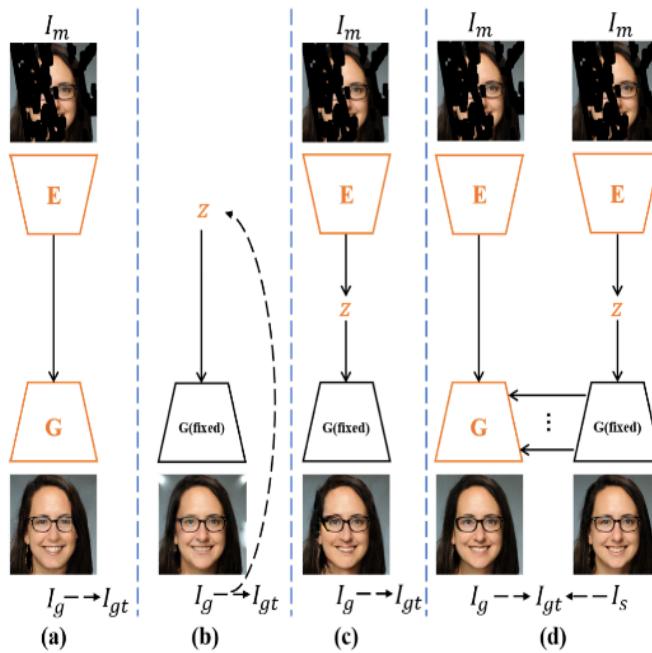
³latent code

⁴Feed-forward network

⁵Feed-forward path

⁶Inversion path

⁷AutoEncoder



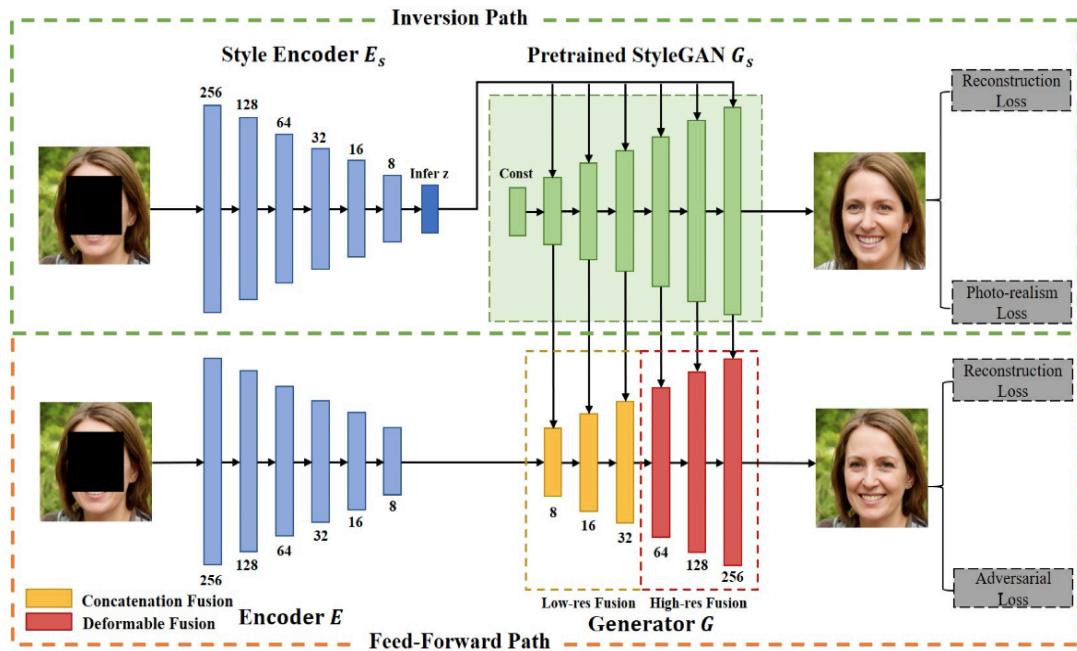
شکل ۴-۱: چهار گونه از روش‌های ترمیم تصویر

قسمت a یک شبکه‌ی ساده‌ی روبه جلو برای ترمیم تصویر را نشان می‌دهد. قسمت b یک روش معکوس سازی شبکه‌های مولد تقابلی مبتنی بر بهینه‌سازی را نشان می‌دهد. قسمت c یک روش معکوس سازی شبکه‌های مولد تقابلی مبتنی بر یادگیری را نشان می‌دهد و قسمت d یک روش دو مسیره را نشان می‌دهد که از معکوس سازی و شبکه‌ی رو به جلو توامان استفاده می‌کند [7].

۲-۴ معماری

۱-۲-۴ مسیر معکوس

در مسیر معکوس همانطور که پیش‌تر هم اشاره شد هدف یافتن نزدیکترین کد پنهان به تصویر تخریب شده و استخراج ویژگی‌های میانی آن از یک شبکه‌ی مولد تقابلی از پیش آموزش دیده است. در این مقاله برای شبکه‌ی مسیر معکوس از شبکه‌ی مولد تقابلی استایل استفاده شده است. ابتدا با رویکرد مبتنی بر یادگیری و استفاده از کدگذار یک کد پنهان z تولید می‌کنیم و بعد این کد پنهان به تک تک لایه‌های تولید کننده‌ی شبکه‌ی مولد تقابلی استایل داده می‌شود تا ویژگی‌های میانی استخراج شوند. خروجی این تولید کننده برای ما اهمیتی ندارد و از آن استفاده نمی‌کنیم چراکه ممکن است کدگذار نتواند نزدیکترین کد پنهان به تصویر را استخراج کند و بنابراین ناهماهنگی در رنگ یا مکان وجود داشته باشد. شکل ۲-۴ ساختار این مسیر را نشان می‌دهد.



شکل ۲-۴: معماری شبکه دومسیره

این شبکه شامل یک مسیر معکوس است که معماری آن در بالای شکل آورده شده است؛ و شامل یک مسیر روبه جلو است که معماری آن در قسمت پایین شکل آورده شده است [۶].

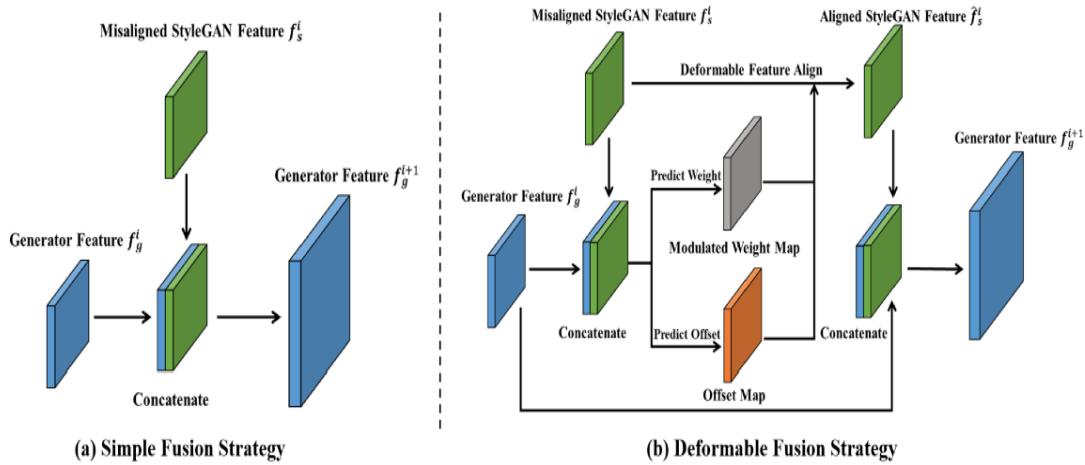
۲-۲-۴ مسیر روبه جلو

در مسیر روبه جلو از یک ساختار خودکدگذار استفاده می‌شود که شامل یک کدگذار و یک تولید کننده است. ابتدا تصویر تخریب شده به کدگذار داده می‌شود تا مجموعه ویژگی کد شده را ایجاد کند سپس این ویژگی به همراه ویژگی‌های میانی شبکه مسیر معکوس به تولید کننده داده می‌شود تا تصویر ترمیم شده را ایجاد کند. شکل ۲-۴ معماری کلی دو مسیر را نشان می‌دهد.

برای ترکیب ویژگی‌های میانی مسیر معکوس و مسیر روبه جلو اولیه می‌تواند الصاق^۸ ویژگی‌های دو مسیر باشد اما با توجه به اینکه امکان خطای عدم تطابق در کد پنهان حاصل از مسیر معکوس وجود دارد، الصاق ویژگی‌ها می‌تواند تخریب و عدم تطابق مکانی ایجاد کند. برای ترکیب ویژگی‌ها در دو مسیر از یک مازول ترکیب شکل پذیر^۹ استفاده می‌شود. با توجه به اینکه اکثر عدم تطابق‌ها در ویژگی‌های سطح بالاتر رخ می‌دهد، بنابراین در سه لایه اولیه شبکه مولد تقابلی صرفاً ویژگی‌های میانی به شبکه می‌تولید کننده مسیر روبه جلو الصاق می‌شوند. اما در سه لایه دوم که ویژگی‌های سطح بالا در خود دارند ابتدا میزان انحراف مکانی ویژگی‌های میان ار تصویر اصلی بیش بینی می‌شود و سپس بر اساس میزان انحراف نقاط متناظر را با یکدیگر منطبق می‌کنیم و سپس آنها را به یکدیگر الصاق می‌کنیم. شکل؟ این موضوع را برای لایه‌های اولیه و ثانویه نشان می‌دهد.

⁸Concatenation

⁹Deformable fuse



شکل ۴-۳: روش ترکیب ویژگی‌های میانی مسیر معکوس در مسیر روبه جلو

برای لایه‌های اولیه که شامل ویژگی‌های فرکانس پایین هستند از الصاقی ویژگی‌ها مطابق قسمت *a* استفاده می‌شود و برای سه لایه‌ی دوم که شامل ویژگی‌های سطح بالاتری هستند از تطابق و الصاق مطابق قسمت *b* استفاده می‌شود [۱۷].

۳-۴ آموزش شبکه

در مسیر معکوس سازی از خطای تقابلی و خطای بازسازی به صورت ترکیبی زیر استفاده می‌شود:

$$\mathcal{L}_{inv} = \mathcal{L}_{sp} + \lambda_{sr}\mathcal{L}_{sr} \quad (1-4)$$

و در مسیر رو به جلو هم به صورت مشابه از خطای بازسازی و خطای تقابلی به صورت زیر استفاده می‌شود.

$$\mathcal{L}_{ff} = \lambda_r\mathcal{L}_r + \mathcal{L}_{adv} \quad (2-4)$$

۴-۴ نتایج

شبکه بر روی دو مجموعه داده‌ی *FFHQ* و *LSUN* آموزش و ارزیابی شده است. مدل از پیش آموزش داده شده‌ای که برای مسیر معکوس استفاده شده مدل *StyleGAN2* است. مقایسه‌ها بر روی تصاویر مجموعه داده با درصد تخریب مختلف و با بررسی ۴ معیار *PSNR*, *SSIM*, *FID* و خطای بازسازی

$L1$ انجام شده است که در همه موارد عملکرد روش پیشنهادی نسبت به روش های پیشین بهتر بوده است. شکل ۴-۴ مجموعه ای از این مقایسه ها با نواحی تخریب بزرگ بر روی تصاویر مجموعه داده $FFHQ$ را نشان می دهد.



شکل ۴-۴: نتایج ترمیم تصویر روش دومسیره روی مجموعه داده $FFHQ$ با توجه به تصاویر ارائه شده در قسمت *a* بخش زیادی از تصاویر تخریب شده است و شبکه به بهترین شکل توانسته آنها را با جزئیات بازسازی نماید [✓].

فصل پنجم

ترمیم تصویر با بهبود محلی و سراسری

۱-۵ ایده‌ی اصلی

در این فصل به چالش‌های موجود در مساله‌ی ترمیم تصویر از جنبه‌ی دیگری می‌نگریم و این جنبه ناحیه‌ی دریافت مدل‌های ترمیم تصویر است. اکثر مدل‌های ترمیم تصویر یک ساختار کدگذار-کدگشا دارند که گاهی همراه با تعدادی اتصالات پرشی است و همین امر سبب می‌شود در بعضی از روش‌ها ناحیه‌ی دریافت حتی از کل تصویر هم بزرگ‌تر باشد. ناحیه‌ی دریافت بزرگ همیشه بهترین گزینه نیست، در بسیاری از تصاویر با ناحیه‌های از دست رفته‌ی شامل بافت‌ها و ساختارهای محلی، استفاده از ناحیه‌ی دریافت بزرگ باعث ایجاد مصنوعات ناخواسته می‌شود. مساله‌ی دیگر این است که ناحیه‌ی دریافت مورد نیاز برای ترمیم تصویر با خرابی‌های مختلف متفاوت است. در این فصل، شبکه‌ای سه مرحله‌ای را معرفی می‌کنیم که با هدف دستکاری ناحیه‌ی دریافت برای بهبود نتایج مطرح شده است. شبکه‌ی ارائه شده از سه زیرشبکه تشکیل شده است. شبکه‌ی اول وظیفه‌ی ترمیم درشت مقیاس^۱ را بر عهده دارد. ساختار اولیه‌ی ناحیه‌ی از دست رفته و برخی از اطلاعات و جزئیات مربوط به بافت را تکمیل می‌کند. شبکه‌ی دوم و سوم که به ترتیب به آنها شبکه‌ی بهبود محلی و شبکه‌ی بهبود سراسری می‌گوییم، با به کارگیری اندازه‌های متفاوتی از ناحیه‌ی دریافت ترمیم اولیه را بهبود می‌دهند. شبکه‌ی بهبود محلی ساختاری کم عمق^۲ دارد که وظیفه‌ی ترمیم ساختارهای محلی و جزئیات بافت را بر اساس نواحی محلی اطراف ناحیه‌ی از دست رفته بر عهده دارد. شبکه‌ی بهبود سراسری نیز با ساختاری مبتنی بر مکانیزم توجه و ناحیه‌ی دریافت بزرگ کیفیت دیداری ناحیه‌ی ترمیم شده را با اطلاعات دورتر و ساختارهای بزرگ‌تر بهبود می‌دهد.

ایده‌ی اولیه‌ی این روش از مقایسه‌ی سه شبکه‌ی یو با ناحیه‌ی دریافت‌های مختلف حاصل شده است. این آزمایش یک شبکه‌ی یو با اندازه‌ی ناحیه‌ی دریافت بزرگ‌تر از تصویر، یک شبکه‌ی یو به همراه یک شبکه‌ی کم عمق با ناحیه‌ی دریافت یک چهارم تصویر و یک مدل با ترکیب دو شبکه‌ی یوی متوالی را در نظر می‌گیرد و خروجی ترمیم تصویر آنها را برای سه نوع تصویر شامل اطلاعات دور و بافت‌های با فاصله و اطلاعات و جزئیات محلی به دست می‌آورد. نتیجه‌ی این آزمایش در شکل ۱-۵ آمده است. این نتیجه بیان می‌کند که شبکه‌ی با ناحیه‌ی دریافت کوچکتر برای ترمیم ساختارهای محلی مناسب‌تر است در حالیکه شبکه‌ی با ناحیه‌ی دریافت بزرگ‌تر برای ترمیم جزئیات دور از دوربین و ساختارهای بزرگ مناسب‌تر است.

¹Coarse inpainting network

²Shallow



شکل ۱-۵: مقایسه‌ی نتایج ترمیم تصویر برای سه شبکه‌ی یو با ناحیه‌ی دریافت‌های مختلف شبکه‌ی یو با ناحیه‌ی بزرگ را با C و شبکه‌ی یو با شبکه‌ی کم عمق با ناحیه‌ی دریافت یک چهارم تصویر را با $C + F_S$ و ترکیب دو شبکه‌ی یو که شبکه‌ی دوم ناحیه‌ی دریافت بزرگی دارد را با $C + F_L$ نشان می‌دهیم [۶].

۲-۵ معماری

۱-۲-۵ شبکه‌ی ترمیم درشت مقیاس

این شبکه ساختار کدگذار-کدگشا دارد که با اتصالات پرشی به هم متصلند. تعداد عملیات‌های زیر نمونه‌برداری و بالانمونه‌برداری آن هشت تاست و ناحیه‌ی دریافت آن 766×766 است که بسیار بزرگتر از تصویر ورودی است. برای کاهش تاری تصویر و افزایش واقع‌نمایی^۳ آن از یک تمایزگر مبتنی بر وصله هم استفاده می‌شود. این تمایزگر تصویر واقعی و تصویر ترمیم شده را می‌گیرد و برای هر وصله‌ای از تصویر واقعی یا جعلی بودن آن را تعیین می‌کند. معماری این شبکه در شکل ۲-۵ آمده است.

³Realism

۲-۲-۵ شبکه‌ی بهبود محلی

برای شبکه‌ی بهبود محلی از شبکه‌ی عمیق کم عمق^۴ استفاده شده است که شامل دو عملیات زیرنمونه‌برداری، بلوک باقیمانده‌ای و دو عملیات بالانمونه‌برداری است. به واسطه‌ی ساختار کم عمق این شبکه ناحیه‌ی دریافت آن 109×109 است که از اندازه‌ی تصویر کوچک‌تر است. با طراحی ارائه شده، نواحی از دست رفته شامل ساختارها و بافت‌های محلی می‌توانند با اطلاعات محلی اطراف ترمیم شوند. معماری این شبکه در شکل ۲-۵ آمده است.

۳-۲-۵ شبکه‌ی بهبود سراسری مبتنی بر مکانیزم توجه

بعد از بهبود محلی تعداد قابل قبولی از مصنوعات دیداری تصویر با به کارگیری اطلاعات محلی حذف می‌شوند؛ اما برخی نواحی از دست رفته با به کارگیری اطلاعات دور از مرز آنها قابلیت ترمیم بهتری دارند. از این رو از یک شبکه‌ی سراسری مبتنی بر مکانیزم توجه برای بهبود آنها بهره می‌بریم. این شبکه اطلاعات کلی و با فاصله‌ی تصویر را به دو طریق دریافت می‌کند: یک، استفاده از ناحیه‌ی دریافت بزرگ و دو، استفاده از مکانیزم توجه. ساختار این شبکه مشابه ساختار شبکه‌ی درشت مقیاس است با این تفاوت که در انتهای کدگذار از سه بلوک توجه استفاده شده است. شکل ۲-۵ معماری این شبکه را نشان می‌دهد.

۳-۵ آموزش شبکه

شبکه‌ی ارائه شده از آموزش پایانه به پایانه^۵ استفاده می‌کند و خطای آموزش نهایی نیز جمع خطاهای سه زیرشبکه به همراه تمایزگر است که در ادامه هر کدام از این خطاهای را معرفی می‌کنیم.

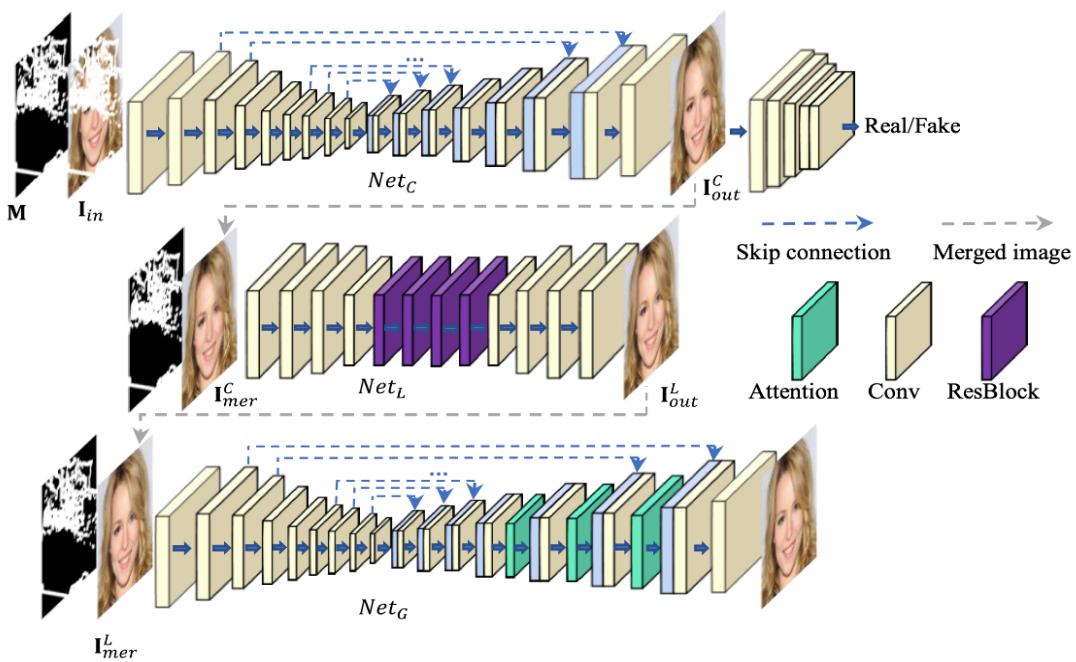
۱-۳-۵ توابع خطای شبکه‌ی درشت مقیاس

شبکه‌ی درشت مقیاس از یک تابع خطای بازسازی پیکسل به پیکسل و خطای تقابلی استفاده می‌کند. خطای بازسازی به صورت خطا روی دو بخش ماسک شده و ماسک نشده محاسبه می‌شود که به صورت زیر می‌باشد:

$$\mathcal{L}_r^C = \mathcal{L}_{valid}^C + \lambda_h \cdot \mathcal{L}_{hole}^C \quad (1-5)$$

⁴Shallow deep network

⁵end-to-end



شکل ۲-۵: معماری کلی شبکه با بهبود محلی و سراسری

شبکه‌ی Net_c شبکه‌ی ترمیم درشت مقیاس را نشان می‌دهد که شامل یک ساختار کدگذار-کدگشا و یک تمايزگر است. شبکه‌ی Net_L شبکه‌ی بهبود محلی را نشان می‌دهد که شامل یک شبکه‌ی کم عمق با بلوک‌های باقی‌ماندهای است. شبکه‌ی Net_G شبکه‌ی بهبود سراسری را نشان می‌دهد که ساختاری مشابه شبکه‌ی درشت مقیاس دارد با این تفاوت که از مکانیزم توجه در لایه‌های کدگشای آن استفاده شده است [۶].

برای خطای تقابلی نیز از خطای کمینه مربعات استفاده می‌شود و بنابراین خطای کلی شبکه درشت مقیاس به صورت زیر است:

$$\mathcal{L}_r^C = \mathcal{L}_{valid}^C + \lambda_h \cdot \mathcal{L}_{hole}^C + \lambda_g \cdot \mathcal{L}_G^C \quad (2-5)$$

۲-۳-۵ توابع خطای شبکه‌ی بهبود محلی

این زیر شبکه از دو خطای بازسازی وزن دار شده با ماسک مشابه شبکه‌ی درشت مقیاس و خطای تغییرات کلی که برای جریمه‌ی هموارسازی به کار گرفته می‌شود استفاده می‌کند. رابطه‌ی خطای کلی به صورت زیر می‌باشد:

$$\begin{aligned} \mathcal{L}_{tv}^L &= \|\mathbf{I}_{mer}^L(i, j+1) - \mathbf{I}_{mer}^L(i, j)\|_1 \\ &+ \|\mathbf{I}_{mer}^L(i+1, j) - \mathbf{I}_{mer}^L(i, j)\|_1. \end{aligned} \quad (3-5)$$

علاوه بر این، زیر شبکه‌ی بهبود محلی مشابه بسیاری از شبکه‌های به کار گرفته شده در ترمیم تصویر از خطای ادراکی و خطای استایل نیز استفاده می‌کند که در محاسبه‌ی هر دو از شبکه‌ی $VGG - 16$ استفاده شده است.

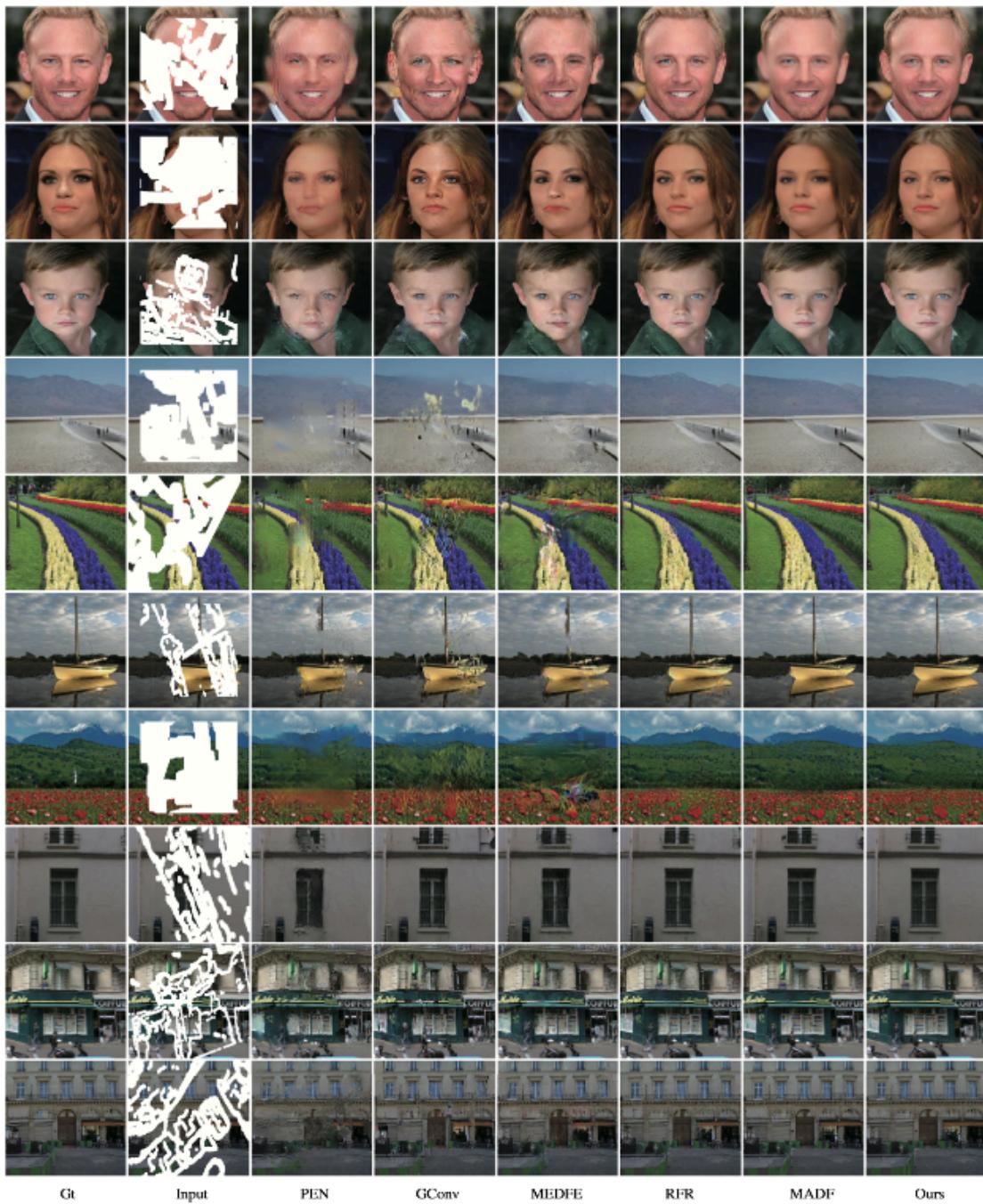
۳-۵ توابع خطای شبکه‌ی بهبود سراسری

در این شبکه نیز مشابه شبکه‌ی محلی از خطای بازسازی وزن دار شده با ماسک، خطای تغییرات کلی، خطای ادراکی و خطای استایل استفاده می‌شود و رابطه‌ی آن به صورت زیر است:

$$\mathcal{L}_L = \mathcal{L}_{valid}^L + \lambda_h \cdot \mathcal{L}_{hole}^L + \lambda_{tv} \cdot \mathcal{L}_{tv}^L + \lambda_{per} \cdot \mathcal{L}_{per}^L + \lambda_{sty} \cdot \mathcal{L}_{sty}^L. \quad (4-5)$$

۴-۵ نتایج

روش ارائه شده روی سه مجموعه داده‌ی *CelebHQ*, *Places2*, *Parisstreetview* و *Parisstreetview* مورد ارزیابی قرار گرفته است. نتایج به دست آمده از این روش بر روی مجموعه داده‌های معرفی شده و در مقایسه با روش‌های کارآمد ترمیم تصویر در درصدهای تخریب مختلف نشان می‌دهد که عملکرد این روش در درصدهای تخریب کم و زیاد نسبت به سایر روش‌ها بهتر است. همچنین شکل ۳-۵ مجموعه‌ای از تصاویر نمونه‌ی این داده‌ها و عملکرد روش در این تصاویر را نشان می‌دهد.



شکل ۳-۵: نتایج ترمیم تصویر شبکه‌ی بهبود محلی و سراسری

نتایج بررسی شده بر روی دو مجموعه داده و تصاویری است که هم شامل اطلاعات دور از دوربین و هم شامل جزئیات محلی مانند اجزای صورت باشند و مقایسه‌ی نتایج این روش نشان می‌دهد که در ترمیم جزئیات نزدیک و محلیو جزئیات دور و سراسری عملکرد مناسبی دارد [۶].

فصل ششم

ترمیم تصویر با جزئیات فرکانس بالا

۱-۶ ایده‌ی اصلی

در این فصل تمرکز بر روی تولید جزئیات واقع‌نمایانه و بافت است. در اکثر شبکه‌های عصبی با پدیده‌ای به نام بایاس طیفی^۱ مواجهیم که بیان می‌کند که یادگیری جزئیات فرکانس بالا برای شبکه‌ها سخت است چرا که شبکه‌ها به سمت یادگیری ویژگی‌های با فرکانس پایین سوگیری یا بایاس دارند. این موضوع به طور خاصل در وظایف بازیابی مانند ترمیم تصویر می‌تواند مشکل‌زا باشد چراکه برای دریافت نتایج واقع‌گرایانه شبکه باید جزئیات فرکانس بالا را تولید کند.

اکثر شبکه‌های متاخر ترمیم تصویر ساختار دو مرحله‌ای دارند؛ به این صورت که در مرحله‌ی اول با استفاده از یک شبکه اطلاعات درشت مقیاس تصویر را تولید می‌کنند و سپس با استفاده از یک شبکه‌ی بهبود، خروجی کلی را بهبود کیفیت می‌دهند و به آن جزئیات می‌افزایند. در این شبکه نیز با توجه به اینکه هدف تولید جزئیات فرکانس بالا است، بعد از تولید کلیات قسمت از دست رفته با یک شبکه‌ی درشت مقیاس، با افزایش رزولوشن تصویر بهبود ثانویه را در یک فضای با رزولوشن بالاتر انجام می‌دهیم. این افزایش رزولوشن به شبکه اجازه می‌دهد که ناپیوستگی‌های محلی را در سطح مناسبتری رفع کند و بایاس طیفی را در رزولوشن نهایی تصویر که کمتر از رزولوشن تصویر در فاز بهبود است کمتر کند. در واقع ایده‌ی اصلی این شبکه اضافه کردن یک مولفه‌ی بالانمونه‌برداری مکعبی^۲ بین شبکه‌ی درشت مقیاس و شبکه‌ی بهبود برای تولید نتایج قبل قبول تر از نظر جزئیات است.

علاوه بر اضافه کردن یک ماثول سوپررزولوشن، ایده‌ی دیگر این شبکه استفاده از یادگیری پیش‌رونده^۳ در فرآیند آموزش شبکه‌ی بهبود است. با توجه به اینکه شبکه‌ی بهبود در رزولوشن بالاتر و با ماسک بزرگتری ترمیم را انجام می‌دهد؛ آموزش آن نسبت به یک شبکه‌ی با رزولوشن پایین تر دشواری بیشتری دارد. در چنین حالتی در یادگیری پیش‌رونده اندازه‌ی ماسک در هر گام یادگیری افزایش پیدا می‌کند. و در نهایت برای تسهیل در آموزش و بهبود جزئیات فرکانس بالا از یکتابع خطای گرادیان نیز استفاده شده است که گرادیان‌های تفاضل بین تصویر ترمیم شده و تصویر اصلی را کمینه می‌کند که یعنی تضمین می‌کند که جزئیات فرکانس بالا تا حد امکان در هر دو مشابه باشند و موجب افزایش تیزی^۴ تصویر بازسازی شده می‌شود.

۲-۶ معماری

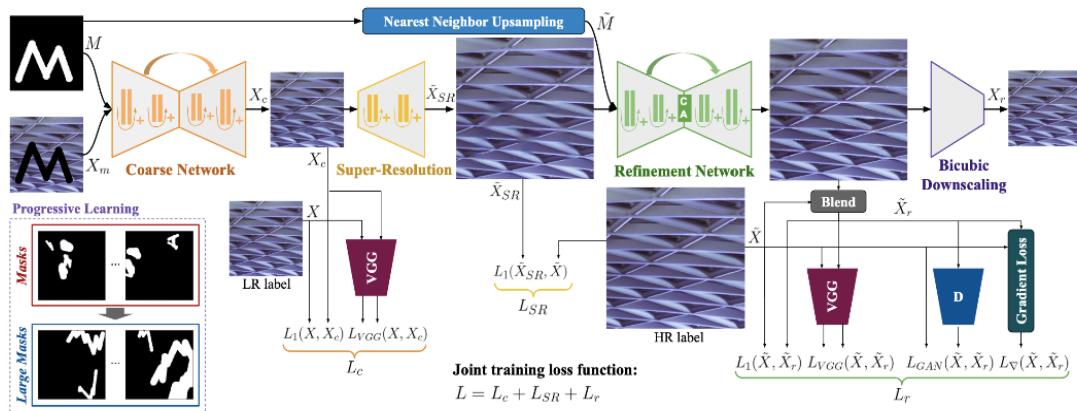
چهارچوب ارائه شده شامل سه شبکه‌ی ترمیم درشت مقیاس، شبکه‌ی سوپررزولوشن و یک شبکه‌ی بهبود رزولوشن بالا می‌باشد. در این ساختار ابتدا هر شبکه به تنها یک پیش‌آموزش داده می‌شود و سپس همه‌ی شبکه‌ها با هم آموزش داده می‌شوند. ساختار کلی این شبکه در شکل^۵ آمده است.

¹Spectral bias

²Bicubic upsampling

³Progressive learning

⁴Sharpness



شکل ۱-۶: معماری شبکه‌ی بزرگنمایی و ترمیم

ایده‌ی اصلی شبکه‌ی بزرگنمایی و ترمیم استفاده از یک مازول سوپر رزوشن بین شبکه‌ی درشت مقیاس و شبکه‌ی بهبود است که در شکل نشان داده شده است [۴].

۱-۲-۶ شبکه‌ی ترمیم درشت مقیاس

وظیفه‌ی این شبکه پرکردن ناحیه‌ی از دست رفته با اطلاعات فرکانس پایین است. در این زیرشبکه از یک ساختار کدگذار کدگشا مبتنی بر کانولوشن‌های دروازه‌ای^۵ مشابه شکل ۲-۶ استفاده شده است [۸]. شبکه‌های کانولوشنی دروازه‌ای از مکانیزم دروازه‌ای^۶ مشابه حافظه‌ی کوتاه مدت بلند^۷ استفاده می‌کنند که وظیفه‌ی آنها کنترل جریان اطلاعات در شبکه با به روز رسانی انتخابی قسمت‌های از نقشه‌ی ویژگی است. در ترمیم تصویر نیز این کانولوشن‌های دروازه‌ای به شبکه کمک می‌کنند تا روی اطلاعات مرتبط تمرکز کند و بر اساس آنها ناحیه‌ی از دست رفته را پر کند.

۲-۲-۶ شبکه‌ی سوپر رزوشن

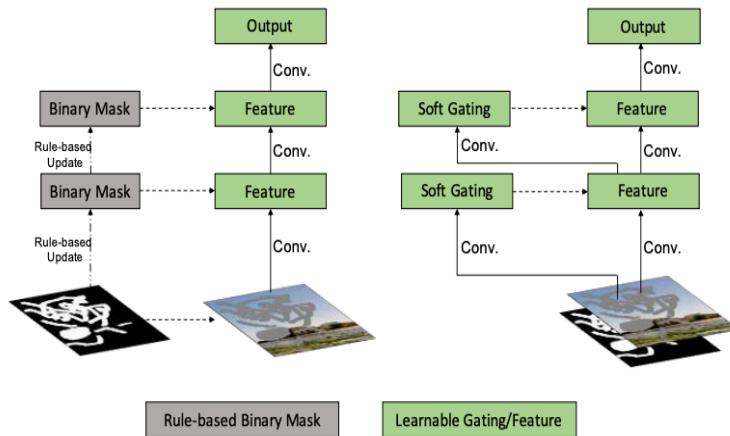
همانطور که در شکل ۱-۶ هم نشان داده شده است، از یک شبکه‌ی سوپر رزوشن برای بزرگ کردن تصویر استفاده می‌کنیم. معماری این شبکه شامل ۴ بلوک باقی‌مانده‌ای و یک لایه‌ی درهم آمیزی پیکسل^۸ است.

⁵Gated convolution

⁶Gating mechanism

⁷long short-term memory(LSTM)

⁸Pixel shuffle layer



شکل ۲-۶: ساختار کانولوشن دروازه‌ای

کانولوشن دروازه‌ای با استفاده از مکانیزم دروازه‌ای جریان اطلاعات در شبکه را کنترل می‌کند. در ترمیم تصویر نیز به شبکه کمک می‌کند تا روی اطلاعات مرتبط تمرکز کند [۸].

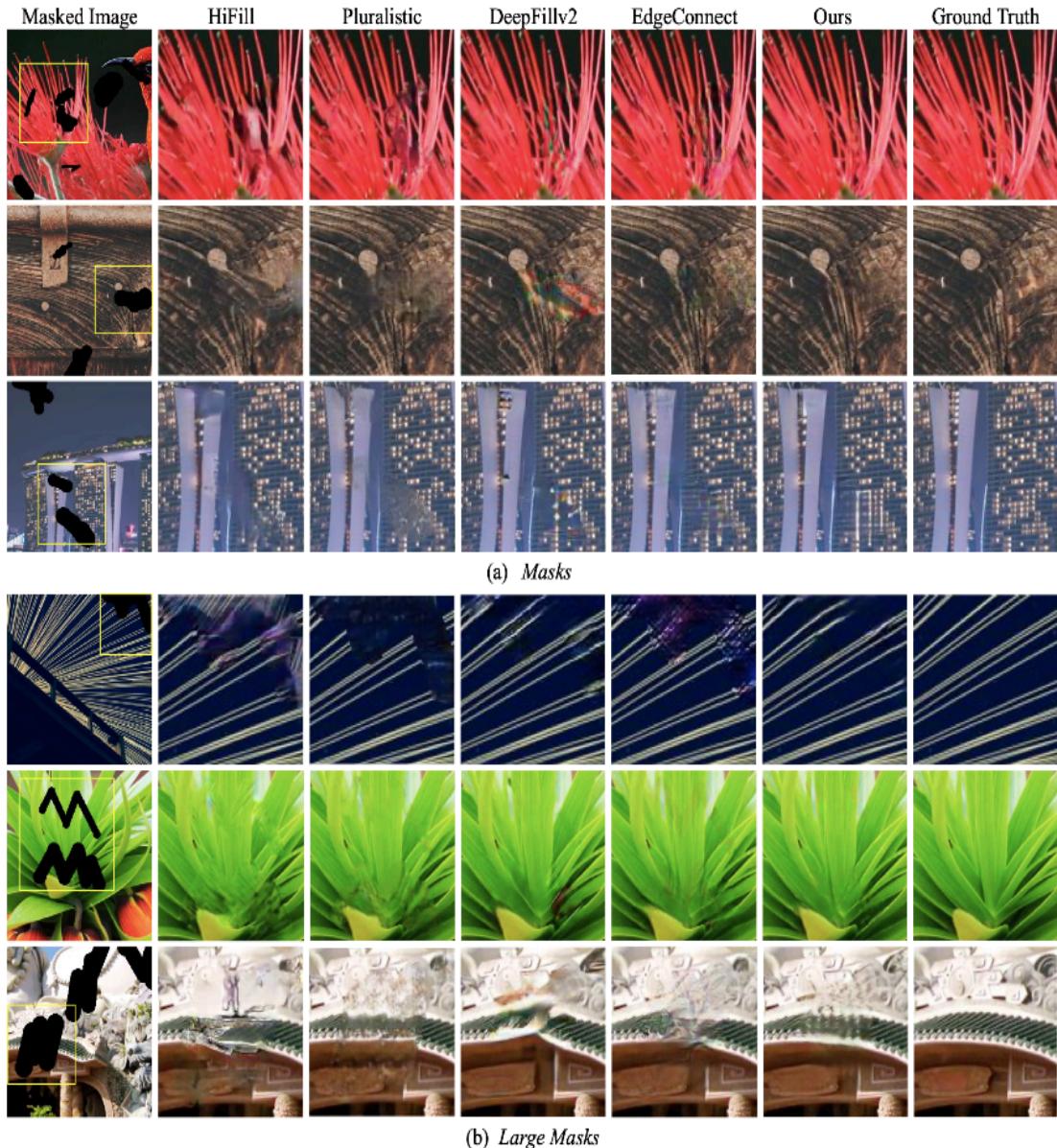
۶-۳-۳ شبکه‌ی بهبود با رزولوشن بالا

این شبکه از تصویر با رزولوشن بالا برای بهبود کمک می‌گیرد تا بتواند اطلاعات ناشی از افزایش رزولوشن را نیز برای بهبود جزئیات به کار گیرد. ساختار این شبکه مشابه یک شبکه‌ی مولد تقابلی وصله‌ای است که این شبکه برای هر وصله از تصویر تشخیص واقعی یا جعلی بودن را می‌دهد. در نهایت پس از خروجی تصویر ترمیم شده از شبکه‌ی بهبود، با به کارگیری یک زیرنمونه‌برداری مکعبی^۹ تصویر به رزولوشن اصلی بازگردانده می‌شود.

۶-۳ نتایج

برای ارزیابی این شبکه از دو مجموعه داده‌ی *Places2* و *Div2k* استفاده شده است که شامل تصاویر 256×256 هستند. شکل ۲-۳-۳ تعدادی از تصاویر این مجموعه داده‌ها و خروجی آنها برای روش ارائه شده و سایر روش‌های موفق ترمیم تصویر را نشان می‌دهد. مهم‌ترین نکته‌ی حائز اهمیت توانایی بازسازی جزئیات در این شبکه در مقایسه با شبکه‌های پیشین است.

⁹Bicubic downsampling



شکل ۳-۶: نتایج ترمیم تصویر با شبکه‌ی بزرگنمایی و ترمیم
نتایج ترمیم تصویر روش بزرگنمایی و ترمیم برای دو اندازه ماسک متفاوت در مقایسه با سایر روش‌های ترمیم
تصویر [۴].

كتاب نامه

- [1] Barnes, Connelly, Shechtman, Eli, Finkelstein, Adam, and Goldman, Dan B. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.*, 28(3):24, 2009.
- [2] Chen, Yuantao, Xia, Runlong, Zou, Ke, and Yang, Kai. Rnon: image inpainting via repair network and optimization network. *International Journal of Machine Learning and Cybernetics*, pages 1–17, 2023.
- [3] Elharrouss, Omar, Almaadeed, Noor, Al-Maadeed, Somaya, and Akbari, Younes. Image inpainting: A review. *Neural Processing Letters*, 51:2007–2028, 2020.
- [4] Kim, Soo Ye, Aberman, Kfir, Kanazawa, Nori, Garg, Rahul, Wadhwa, Neal, Chang, Huiwen, Karnad, Nikhil, Kim, Munchurl, and Liba, Orly. Zoom-to-inpaint: Image inpainting with high-frequency details. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 477–487, 2022.
- [5] Pathak, Deepak, Krahenbuhl, Philipp, Donahue, Jeff, Darrell, Trevor, and Efros, Alexei A. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.
- [6] Quan, Weize, Zhang, Ruisong, Zhang, Yong, Li, Zhifeng, Wang, Jue, and Yan, Dong-Ming. Image inpainting with local and global refinement. *IEEE Transactions on Image Processing*, 31:2405–2420, 2022.

- [7] Wang, Wentao, Niu, Li, Zhang, Jianfu, Yang, Xue, and Zhang, Liqing. Dual-path image inpainting with auxiliary gan inversion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 11421–11430, 2022.
- [8] Yu, Jiahui, Lin, Zhe, Yang, Jimei, Shen, Xiaohui, Lu, Xin, and Huang, Thomas S. Free-form image inpainting with gated convolution. In Proceedings of the IEEE/CVF international conference on computer vision, pages 4471–4480, 2019.
- [9] Zeng, Yanhong, Fu, Jianlong, Chao, Hongyang, and Guo, Baining. Learning pyramid-context encoder network for high-quality image inpainting. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 1486–1494, 2019.
- [10] Zeng, Yanhong, Fu, Jianlong, Chao, Hongyang, and Guo, Baining. Aggregated contextual transformations for high-resolution image inpainting. IEEE Transactions on Visualization and Computer Graphics, 2022.