



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده ریاضی و علوم کامپیوتر

پروژه کارشناسی
علوم کامپیوتر

تنظیم فرایارامترهای شبکه های بازگشایی عمیق برای بهبود کیفیت تصاویر

نگارش
پوریاعلیمرادپور

استاد راهنما
دکتر فاطمه شاکری

استاد مشاور
دکتر بهزاد نجفی

زمستان ۱۴۰۲

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

صفحه فرم ارزیابی و تصویب پایان نامه - فرم تأیید اعضاء کمیته دفاع

در این صفحه فرم دفاع یا تایید و تصویب پایان نامه موسوم به فرم کمیته دفاع - موجود در پرونده آموزشی - را قرار دهید.



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

به نام خدا

تعهدنامه اصالت اثر

تاریخ: زمستان ۱۴۰۲

اینجانب **پوریا علیمیرادپور** متعهد می‌شوم که مطالب مندرج در این پایان‌نامه حاصل کار پژوهشی اینجانب تحت نظارت و راهنمایی اساتید دانشگاه صنعتی امیرکبیر بوده و به دستاوردهای دیگران که در این پژوهش از آنها استفاده شده است مطابق مقررات و روال متعارف ارجاع و در فهرست منابع و مآخذ ذکر گردیده است. این پایان‌نامه قبلاً برای احراز هیچ مدرک هم‌سطح یا بالاتر ارائه نگردیده است. در صورت اثبات تخلف در هر زمان، مدرک تحصیلی صادر شده توسط دانشگاه از درجه اعتبار ساقط بوده و دانشگاه حق پیگیری قانونی خواهد داشت.

کلیه نتایج و حقوق حاصل از این پایان‌نامه متعلق به دانشگاه صنعتی امیرکبیر می‌باشد. هرگونه استفاده از نتایج علمی و عملی، واگذاری اطلاعات به دیگران یا چاپ و تکثیر، نسخه‌برداری، ترجمه و اقتباس از این پایان‌نامه بدون موافقت کتبی دانشگاه صنعتی امیرکبیر ممنوع است. نقل مطالب با ذکر مآخذ بلامانع است.

پوریا علیمیرادپور

امضا

سپاسگزاری

با کمال محبت و قدردانی، این پایان نامه را تقدیم می کنم به کسانی که با حمایت های بی دریغ و راهنمایی های ارزشمند خود، مرا در این سفر علمی یاری نمودند.

از استاد راهنمای گران قدرم، خانم دکتر فاطمه شاکری، که با صبر و دانش خود، مرا در طول این مسیر هدایت فرمودند، صمیمانه سپاسگزارم. همچنین، از دانشجوی دکتری ایشان، خانم سپیده خرمایی پور، که با اشتراک گذاری دانش و تجربیات ارزنده خود، به من کمک کردند تا به درک عمیق تری از موضوع تحقیق دست یابم، قدردانی می کنم.

از استاد مشاور دوره کارشناسی، آقای دکتر بهزاد نجفی، که با نگاه دقیق و انتقادات سازنده خود، مرا در ارتقاء کیفیت کارم یاری رساندند، تشکر ویژه دارم. این پایان نامه حاصل تلاش های مشترک و همفکری های بیشائبه ای است که در کنار این بزرگواران تجربه کردم. امیدوارم که نتایج این تحقیق، گامی مثبت در جهت پیش برد دانش و فناوری باشد.

پایان علم را دور
زمستان ۱۴۰۲

چکیده

بهبود کیفیت تصاویر، چالشی مهم در پردازش تصویر است. روش‌های سنتی مانند روش‌های مبتنی بر مدل و مبتنی بر متغیر، با وجود مزایایی که دارند، محدودیت‌هایی نیز دارند. در سال‌های اخیر، شبکه‌های بازگشایی عمیق به عنوان رویکردی نوین برای حل مسائل بد وضع معکوس در بهبود کیفیت تصاویر، توجه بسیاری را به خود جلب کرده‌اند. این شبکه‌ها با ترکیب مزایای روش‌های مبتنی بر مدل و یادگیری عمیق (شبکه‌های عصبی عمیق)، کارایی را حفظ می‌کنند و تفسیرپذیری یادگیری عمیق را بهبود می‌بخشند. در این مقاله، مروری جامع بر معماری‌های مختلف شبکه‌های بازگشایی عمیق برای بهبود کیفیت تصاویر ارائه می‌شود. ابتدا، به معرفی مفهوم بازگشایی عمیق و مزایای آن نسبت به روش‌های سنتی می‌پردازیم. سپس، به بررسی معماری‌های مختلف شبکه‌های بازگشایی عمیق می‌پردازیم.

در آخر، به طور مختصر به کاربردهای این شبکه‌ها در زمینه‌های مختلف بهبود کیفیت تصاویر، مانند رفع تاری تصاویر، بهبود تصاویر کم‌نور، ترمیم تصاویر، افزایش وضوح تصاویر و رفع نویز تصاویر، اشاره می‌کنیم.

واژه‌های کلیدی:

(شبکه‌های بازگشایی عمیق، شبکه عصبی، بهبود کیفیت تصاویر، مسائل بد وضع معکوس)

فهرست مطالب

صفحه

عنوان

۲	۱ مقدمه
۵	۲ شبکه‌های بازگشایی عمیق
۶	۱-۲ رفع تاری تصاویر (Image deblurring)
۸	۱-۱-۲ الگوریتم نقطه داخلی نزدیک (Proximal IPM)
۱۰	۲-۱-۲ معماری شبکه iRestNet
۱۲	۲-۲ بهبود تصاویر کم نور (Low-light image enhancement)
۱۴	۱-۲-۲ الگوریتم متنی بر Retinex برای بهبود کیفیت تصاویر کم نور
۱۶	۲-۲-۲ معماری شبکه URetinex-Net
۱۸	۳-۲ ترمیم تصاویر (Image restoration)
۲۰	۱-۳-۲ الگوریتم سنتی Proximal Gradient Descent
۲۱	۲-۳-۲ معماری شبکه DGUNet
۲۵	۴-۲ افزایش وضوح تصاویر (Image super-resolution)
۲۸	۱-۴-۲ الگوریتم منعطف SISR
۲۹	۲-۴-۲ معماری شبکه USRNet
۳۳	۵-۲ رفع نویز تصاویر (Image denoising)
۳۵	۱-۵-۲ مازول حذف نویز مبتنی بر ویژگی (FDM)
۳۷	۲-۵-۲ معماری شبکه DUMRN
۳۸	۳ نتیجه‌گیری
۳۹	۱-۳ جدول مقایسه نهایی
۴۱	منابع و مراجع

فصل اول

مقدمه

در سال‌های اخیر، با پیشرفت فناوری در بسیاری از حوزه‌ها از جمله پزشکی، علوم مهندسی و هنر، نیاز به تصاویر با کیفیت بالا افزایش یافته است؛ در ارتباط با مبحث بهبود کیفیت تصاویر، بازسازی تصویر (Image Restoration یا IR) به دنبال بازبازی تصویر با کیفیت x از نسخه‌ی تخریب شده‌ی آن y است. فرآیند تخریب در ریاضیات به طور کلی به صورت زیر فرموله می‌شود:

$$y = Ax + n \quad (1-1)$$

که A ماتریس مخرب، و n نماد نویز افزوده شده است. به این نوع مسائل، مسائل بد وضع معکوس (Ill-posed Inverse Problems) گفته می‌شود.

روش‌های گوناگونی برای حل مسائل بد وضع معکوس در بهبود کیفیت تصاویر تا به امروز ارائه شده. از جمله این روش‌ها می‌توان به روش‌های مبتنی بر مدل (model-based methods)، روش‌های مبتنی بر تنوع یا همان مبتنی بر متغیر (variational-based methods) و روش‌های یادگیری عمیق (deep-learning methods) اشاره کرد.

روش‌های مبتنی بر متغیر به طور گسترده برای مسائل بد وضع معکوس ارائه شده استفاده می‌شوند زیرا آنها توانایی تعبیه دانش قبلی در مورد راه حل را دارند. با این حال، سطح عملکرد این روش‌ها به طور قابل توجهی به مجموعه‌ای از پارامترها بستگی دارد که می‌توان آن‌ها را از طریق روش‌های محاسباتی پرهزینه و زمان‌بر تخمین زد. همچنین روش‌های مبتنی بر مدل معمولاً تصاویر تمیز پنهان را با حل مسائل بهینه‌سازی پیچیده بازسازی می‌کنند که کاربرد عملی آن‌ها را محدود می‌کند. در مقابل، یادگیری عمیق معماری‌های بسیار عمومی و کارآمدی را ارائه می‌دهند، البته به قیمت توضیح‌پذیری سخت، زیرا اغلب به عنوان یک جعبه سیاه طراحی شده‌اند که شفافیت و قابلیت تفسیر ندارند.

بازگشایی عمیق یک رویکرد راحت برای ترکیب رویکردهای مبتنی بر متغیر و یادگیری عمیق ارائه می‌کند. همچنین با بهره‌گیری از روش‌های مبتنی بر مدل و یادگیری عمیق، بسیاری از روش‌های بازگشایی عمیق، روش‌های بهینه‌سازی استاندارد را در شبکه‌های عصبی عمیق ترکیب کرده و با این کار کارایی را حفظ می‌کنند و تفسیرپذیری یادگیری عمیق را بهبود می‌بخشند.

وابسته به تعریف A و n در فرمول 1-1، بازسازی تصویر می‌تواند به زیر مسئله‌هایی مانند رفع تاری تصاویر (Image deblurring) [1]، بهبود تصاویر کم نور (Low-light image enhancement) [2]، ترمیم تصاویر (Image restoration) [3]، افزایش وضوح تصاویر (Image super-resolution) [4]، افزایش وضوح دقیق و سبک تصاویر (Accurate and lightweight image super-resolution) [5]، رفع نویز تصاویر

(Image denoising) [۶]، گسترش یابد که به ترتیب باعث ابداع معماری شبکه‌های بازگشایی عمیق iRestNet^۱، URetinex-Net^۲، DGUNet^۳، USRNet^۴، MoG-DUN^۵ و DUMRN^۶ شده‌اند.

در این مقاله، مروری بر برخی معماری شبکه‌های بازگشایی عمیق مختلف نام برده شده خواهیم داشت و با نحوه کار هریک بیشتر آشنا خواهیم شد. همچنین کد منبع این پروژه حاوی تمامی معماری های نام برده شده همراه با دیتاست‌های استفاده شده، در گیت‌هاب به آدرس زیر موجود است:

<https://github.com/poripouria/Bachelors-Project/>

^۱ لینک گیت‌هاب پروژه: <https://github.com/mccorbineau/iRestNet>

^۲ لینک گیت‌هاب پروژه: <https://github.com/AndersonYong/URetinex-Net>

^۳ لینک گیت‌هاب پروژه: <https://github.com/MC-E/DGUNet>

^۴ لینک گیت‌هاب پروژه: <https://github.com/cszn/USRNet>

^۵ لینک گیت‌هاب پروژه: <https://github.com/QianNing0/MoG-DUN>

^۶ لینک گیت‌هاب پروژه: <https://github.com/Xujz19/DUMRN>

فصل دوم

شبکه‌های بازگشایی عمیق

همانطور که در فصل قبل اشاره شد، بازگشایی عمیق به عنوان ترکیبی از روش‌های مبتنی بر متغیر و مبتنی بر مدل با روش‌های یادگیری عمیق، مزایای هر دو بخش را به ارمغان می‌آورد. از یک طرف، کارایی روش‌های مبتنی بر متغیر و مبتنی بر مدل را حفظ می‌کند و از طرف دیگر، تفسیرپذیری روش‌های یادگیری عمیق را بهبود می‌بخشد.

در این فصل، به معماری‌های مختلف شبکه‌های بازگشایی عمیق مختلف خواهیم پرداخت و نحوه عملکرد هر یک را بررسی خواهیم کرد.

۱-۲ رفع تاری تصاویر (Image deblurring)

با شروع از یک فرمول‌بندی متغیر^۱ برای بازیابی تصویر، نویسندگان مقاله "Deep Unfolding of a Proximal Interior Point Method for Image Restoration" [۱]، iRestNet، یک معماری شبکه عصبی را توسعه داده‌اند که با بازگشایی یک الگوریتم نقطه داخلی نزدیک^۲ (Proximal IPM) به دست می‌آید. محدودیت‌های الزامی، کدگذاری ویژگی‌های مطلوب برای تصویر بازیابی شده، به لطف یک مانع لگاریتمی^۳ در شبکه گنجانده شده‌اند، در حالی که پارامتر مانع^۴، اندازه قدم^۵ و وزن جریمه^۶ توسط شبکه یادگرفته می‌شود. برای محدودیت‌های مختلف، بیان دقیقی برای گرادیان عملگر نزدیکی ارائه شده است. این امر امکان آموزش شبکه iRestNet را با استفاده از روش گرادیان کاهشی و پس انتشار^۷ فراهم می‌کند. علاوه بر این، نویسندگان مقاله نتایج نظری را در رابطه با پایداری شبکه برای مسائل معکوس رایج ارائه می‌کنند. آزمایش‌های عددی روی مسائل رفع تاری تصویر نشان می‌دهد که رویکرد پیشنهادی از نظر کیفیت تصویر با روش‌های پیشرفته متغیر و یادگیری ماشین مقایسه مطلوبی دارد.

به منظور یافتن راه‌حل مناسب برای یک مسئله معکوس بد وضع ارائه شده مانند ۱-۱، روش‌های متغیر پیش‌فرض‌ها (اطلاعات) قبلی را در مورد متغیر مطلوب x ، از طریق محدودیت‌ها^۸ یا توابع منظم‌سازی^۹،

^۱ variational

^۲ proximal interior point method

^۳ logarithmic barrier

^۴ barrier parameter

^۵ stepsize

^۶ penalization weight

^۷ backpropagation

^۸ constraints

^۹ regularization functions

ترکیب می‌کنند. این منجر به مسئله کمینه‌سازی زیر می‌شود:

$$\min_{x \in C} f(Hx, y) + \lambda \mathcal{R}(x) \quad (۱-۲)$$

که در آن $f: \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}$ یک تابع وفاداری داده^{۱۰} است که با توجه به اولین متغیر خود محدب است و مربوط به مدل تخریب است، $\mathcal{R}: \mathbb{R}^n \rightarrow \mathbb{R}$ یک تابع منظم‌سازی محدب است، $\lambda \in]0, +\infty[$ یک پارامتر منظم‌سازی و C زیر مجموعه‌ای از \mathbb{R}^n است. اگرچه این رویکرد مفید است، اما گاهی اوقات به دلیل پیچیدگی آن محدود می‌شود: حل ۱-۲ ممکن است به الگوریتم‌های پیشرفته‌ای نیاز داشته باشد که ممکن است برای برنامه‌های بلادرنگ^{۱۱} بسیار کند باشد. علاوه بر این، λ پارامتری است که باید تنظیم شود و \mathcal{R} معمولاً با یک یا چند پارامتر، پارامتری می‌شود که انتخاب بهینه آن ممکن است به شدت به داده‌های موجود بستگی داشته باشد. این پارامترها اغلب به صورت دستی تنظیم می‌شوند با این حال، این روش‌ها اغلب زمان بر هستند و موفقیت آن‌ها همیشه تضمین شده نیست. علاوه بر این، با وجود تلاش‌های متعدد در طراحی مدل‌های پیچیده، راه‌حل ۱-۲ می‌تواند دورتر از x نسبت به یک تکرار میانی تولید شده توسط یک الگوریتم معین که برای حل ۱-۲ استفاده می‌شود باشد. چنین پدیده‌ای توسعه روش‌های توقف اولیه را توجیه می‌کند، جایی که روش تکراری قبل از همگرایی متوقف می‌شود. یافتن زمان توقف بهینه به الگوریتم بستگی دارد که ممکن است تشریح کند که چرا این تکنیک‌ها در حال حاضر به توابع هزینه نسبتاً ساده محدود می‌شوند.

شبکه‌های عصبی عمیق^{۱۲} (DNN)، و به‌ویژه شبکه‌های عصبی پیچشی^{۱۳} (CNN)، عملکرد خوبی را برای کاربردهای مختلف مرتبط با مسائل معکوس ارائه می‌دهند. شبکه‌های عصبی عمیق برای مسائل معکوس اغلب با یک مرحله پیش‌پردازش شروع می‌شوند. در واقع، تخمین تقریبی x را می‌توان با استفاده از معکوس یا شبه‌معکوس ماتریس H پیدا کرد. از این رو، در این زمینه، DNN‌ها به عنوان حذف‌کننده نویز و مصنوعات استفاده می‌شوند. با این حال، از آنجایی که پیش‌فرض‌های (دانش) قبلی در مورد خروجی آن به سختی می‌تواند در یک DNN که در بیشتر موارد به عنوان یک جعبه سیاه^{۱۴} در نظر گرفته می‌شود، گنجانده شود، توضیح‌پذیری و قابلیت اطمینان چنین روش‌هایی می‌تواند مورد تردید قرار

^{۱۰} data-fidelity function

^{۱۱} real-time

^{۱۲} deep neural networks

^{۱۳} convolutional neural networks

^{۱۴} black-box

گیرد. علاوه بر این، مرحله پیش‌پردازش، به خودی خود، می‌تواند شامل جریمه باشد، بنابراین به حل مسئله به فرم ۱-۲ می‌رسد، که در آن وزن منظم‌ساز به شدت به سطح نویز بستگی دارد. این واقعیت جالب توجه که این رویکرد از تعداد محدودی لایه استفاده می‌کند، می‌تواند به عنوان الگو در روش‌های توقف اولیه در نظر گرفته شود. البته شایان ذکر است که در الگوریتم‌های بازگشایی، تعداد تکرارها (یعنی لایه‌ها) در مرحله آموزش آفلاین تنظیم می‌شود و سپس برای همه تصاویر آزمایشی ثابت می‌شود، که با استراتژی‌های توقف اولیه که در آن تعداد تکرار معمولاً برای هر تصویر پردازش شده فرق دارد، متفاوت است.

در این مقاله، نویسندگان یک معماری شبکه عصبی جدید به نام iRestNet را پیشنهاد می‌کنند که با باز کردن یک الگوریتم نقطه داخلی نزدیک در تعداد محدودی از تکرارها به دست می‌آید. یکی از ویژگی‌های کلیدی این الگوریتم این است که به لطف یک مانع لگاریتمی، فقط تکرارهای عملی را تولید می‌کند. این مانع دانش قبلی را قادر می‌سازد تا مستقیماً در iRestNet گنجانده شود و برخلاف انداختن تصویر^{۱۵} بر روی C ، روش گرادیان کاهشی و پس‌انتشار را در سراسر شبکه امکان‌پذیر می‌سازد. از این رو می‌توان از گرادیان کاهشی برای آموزش استفاده کرد. اندازه گام^{۱۶}، پارامتر مانع^{۱۷}، و وزن منظم‌سازی^{۱۸} در سراسر شبکه بازگشایی و برای هر لایه آموخته می‌شود. بنابراین، هنگامی که شبکه آموزش داده شد، کارکرد آن بر روی تصاویر آزمایشی تنها به زمان اجرای کوتاهی برای هر تصویر بدون جستجوی پارامتر، بر خلاف روش‌های سنتی متغیر، نیاز دارد.

۱-۱-۲ الگوریتم نقطه داخلی نزدیک (Proximal IPM)

به طور کلی، مسئله ۱-۲ به دلیل محدودیت‌های نامساوی، حتی برای منظم‌سازی‌های ساده، راه‌حل بسته ندارد، بنابراین باید از یک حل‌کننده تکراری استفاده شود. روش‌های استاندارد نقطه داخلی نیاز به معکوس کردن چندین دستگاه خطی $n \times n$ دارند، که منجر به پیچیدگی محاسباتی بالا برای مسائل در مقیاس بزرگ می‌شود. با این وجود، اخیراً نشان داده شده است که ترکیب چارچوب نقطه داخلی^{۱۹} با یک استراتژی جلو-عقب تقریبی^{۲۰} منجر به حل‌کننده‌های بسیار رقابتی برای مسائل معکوس می‌شود.

^{۱۵} projection

^{۱۶} stepsize

^{۱۷} barrier parameter

^{۱۸} regularization weight

^{۱۹} interior point framework

^{۲۰} proximal forward-backward strategy

ایده پشت IPM ها این است که مسئله بهینه‌سازی محدود اولیه را با دنباله‌ای از مسائل فرعی بدون محدودیت به شکل زیر جایگزین کنیم:

$$\min_{x \in \mathbb{R}^n} f(Hx, y) + \lambda \mathcal{R}(x) + \mu \mathcal{B}(x) \quad (2-2)$$

که در آن \mathcal{B} تابع مانع لگاریتمی با مشتق نامحدود در مرز دامنه شدنی^{۲۱} است:

$$(\forall x \in \mathbb{R}^p) \quad \mathcal{B}(x) = \begin{cases} -\sum_{i=1}^p \ln(c_i(x)) & x \in \text{int}C \\ +\infty & \text{otherwise} \end{cases} \quad (3-2)$$

و $\mu \in]^\circ, +\infty[$ به اصطلاح پارامتر مانع است که در طول فرآیند کمینه‌سازی ناپدید می شود. ما فرض کردیم که یا $f(Hx, y) + \lambda \mathcal{R}$ اجباری است، یا C محدود است، بنابراین، مجموعه راه حل‌های ۱-۲ محدود است. از آنجایی که $\text{int}C$ (تعریف شده در بخش ۱.۲ مقاله [۱]) تهی نیست، وجود راه حل برای ۳-۲ تضمین شده است.

به لطف عملگر نزدیکی^{۲۲}، IPM نیازی به وارونگی ماتریسی ندارد. هنگامی که عملگر نزدیکی به روشی دقیق محاسبه می‌شود، IPM پیشنهادی را می‌توان به‌عنوان الگوریتم ۱ (شکل ۱-۲ را ببینید) بازنویسی کرد که هم‌گرایی آن تحت برخی مفروضات ثابت شده است. شایان ذکر است که الگوریتم ۲ فقط به محاسبه عملگر نزدیکی مانع لگاریتمی نیاز دارد.

برای تمامی $g \in \Gamma_\circ(\mathbb{R}^n)$ و $\gamma \in]^\circ, +\infty[$ و $x \in \mathbb{R}^n$ عملگر نزدیکی γg در x به صورت زیر تعریف می‌شود:

$$\text{prox}_{\gamma g}(x) = \arg \min_{u \in \mathbb{R}^n} \frac{1}{\gamma} \|x - u\|_\gamma^2 + \gamma g(u). \quad (4-2)$$

که $\Gamma_\circ(\mathbb{R}^q)$ مجموعه‌ای از توابع را نشان می‌دهد که مقادیر $\mathbb{R} \cup \{+\infty\}$ را می‌گیرند و مناسب، محدب و شبه‌پیوسته بر روی \mathbb{R}^q هستند.

^{۲۱} feasible

^{۲۲} proximity operator

Algorithm 1 Exact version of the proximal IPM in [50] applied to problem (2).

Let $x_0 \in \text{int}\mathcal{C}$, $\underline{\gamma} > 0$ and $(\gamma_k)_{k \in \mathbb{N}}$ be a sequence such that $(\forall k \in \mathbb{N}) \underline{\gamma} \leq \gamma_k$;
for $k = 0, 1, \dots$ **do**
 $x_{k+1} = \text{prox}_{\gamma_k(h(\cdot, y, \lambda) + \mu_k \mathcal{B})}(x_k)$
end for

Algorithm 1 requires evaluating the proximity operator of the sum of the barrier and the regularized cost function, which can be an issue since, in most of the cases, this operator does not have a closed-form expression. This is the reason why we propose to modify it by introducing a forward step, which leads to Algorithm 2.

Algorithm 2 Proposed forward-backward proximal IPM.

Let $x_0 \in \text{int}\mathcal{C}$, $\underline{\gamma} > 0$ and $(\gamma_k)_{k \in \mathbb{N}}$ be a sequence such that $(\forall k \in \mathbb{N}) \underline{\gamma} \leq \gamma_k$;
for $k = 0, 1, \dots$ **do**
 $x_{k+1} = \text{prox}_{\gamma_k \mu_k \mathcal{B}}(x_k - \gamma_k \nabla_1 h(x_k, y, \lambda))$
end for

شکل ۱-۲

۲-۱-۲ معماری شبکه iRestNet

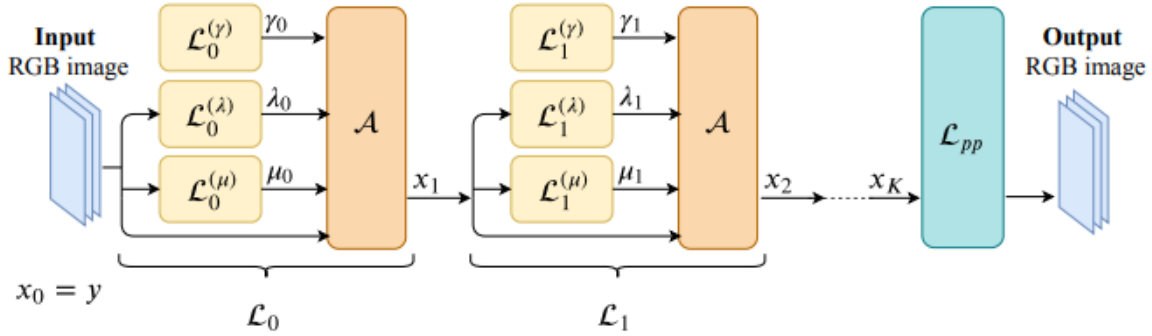
پیشنهاد نویسندگان مقاله اتخاذ یک استراتژی یادگیری با ناظر از مجموعه آموزشی از تصاویر، به منظور تعیین تنظیمات بهینه برای پارامترهای الگوریتم ۲ است که باید به کیفیت بازیابی تصویر بهینه منجر شود. برای این هدف، الگوریتم ۲ بر روی K تکرار بازگشایی می‌شود و پارامتر منظم‌سازی λ در سراسر شبکه یکپارچه می‌شود، تا انعطاف‌پذیری بیشتری فراهم شود. قانون بروزرسانی در یک تکرار معین $k \in \{0, \dots, K-1\}$ به صورت زیر خواهد بود:

$$x_{k+1} = \mathcal{A}(x_k, \mu_k, \gamma_k, \lambda_k) \quad (5-2)$$

$$\mathcal{A}(x_k, \mu_k, \gamma_k, \lambda_k) = \text{prox}_{\gamma_k \mu_k \mathcal{B}}(x_k - \gamma_k \nabla_1 h(x_k, y, \lambda)) \quad (6-2)$$

برای هر $k \in \{0, \dots, K-1\}$ ، ما k -امین لایه \mathcal{L}_k را به عنوان تداعی سه ساختار پنهان، $\mathcal{L}_k^{(\mu)}$ ، $\mathcal{L}_k^{(\gamma)}$ و $\mathcal{L}_k^{(\lambda)}$ می‌سازیم، که به دنبالش به‌روزرسانی \mathcal{A} را دارد. ساختارهای $\mathcal{L}_k^{(\mu)}$ ، $\mathcal{L}_k^{(\gamma)}$ و $\mathcal{L}_k^{(\lambda)}$ به ترتیب استنباط پارامتر مانع μ_k ، اندازه گام γ_k و وزن منظم‌ساز λ_k را دنبال می‌کنند. از آنجایی که تعداد محدود K از لایه‌ها (بروزرسانی‌ها) استفاده می‌شود، همگرایی طرح حاصل مشکلی ندارد. توجه داشته باشید که نویسندگان مقاله همچنین در چارچوب خود اجازه استفاده از یک مرحله پس‌پردازش را پس از عبور

از K لایه می‌دهند که با \mathcal{L}_{pp} نشان داده می‌شود. معماری حاصل در شکل ۲-۲ نشان داده شده است.



شکل ۲-۲: معماری سراسری شبکه iRestNet

که به طور خلاصه $\mathcal{L}_k^{(\gamma)} = \text{Softplus}(a_k)$ که یک پارامتر اسکالر از شبکه است که هنگام آموزش یادگرفته می‌شود. دلیل استفاده از Softplus این است که مسئله‌ی مرگ ReLU را ندارد و به صورت $\text{Softplus}(z) = \ln(1 + \exp(z))$ ($\forall z \in \mathbb{R}$) تعریف می‌شود. همچنین تعریف $\mathcal{L}_k^{(\lambda)}$ به نوع مسئله وابسته است و در شکل زیر معماری $\mathcal{L}_k^{(\mu)}$ و \mathcal{L}_{pp} نمایش داده شده است.

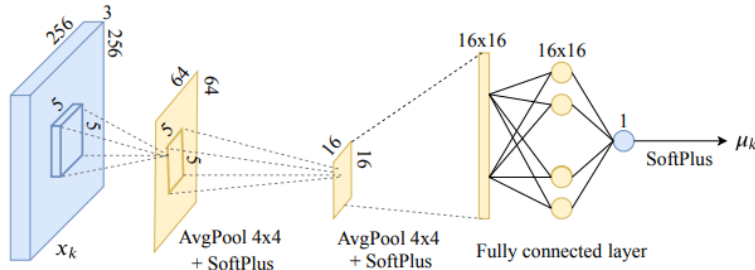


Figure 3. Architecture of $\mathcal{L}_k^{(\mu)}$.

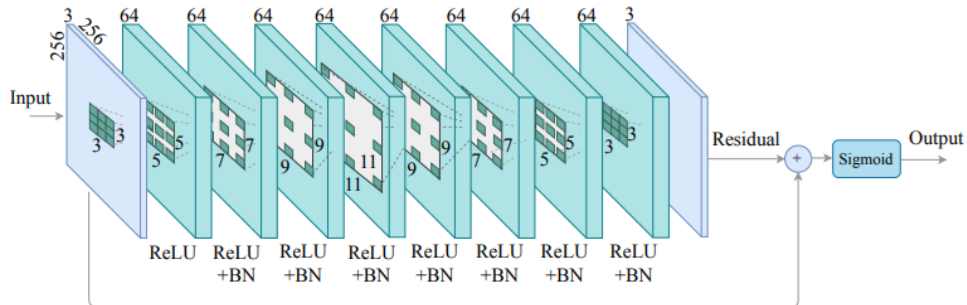


Figure 4. Architecture of \mathcal{L}_{pp} . BN: batch normalization.

۲-۲ بهبود تصاویر کم نور (Low-light image enhancement)

روش‌های مبتنی بر مدل رتینکس^{۲۳} نشان داده‌اند که با استفاده از پیش‌فرض‌های به‌خوبی طراحی شده در دستکاری لایه به لایه برای افزایش کیفیت تصاویر کم‌نور، مؤثر هستند. با این حال، دانش پیشین دست‌ساز رایج و راه‌حل‌های مبتنی بر بهینه‌سازی منجر به عدم سازگاری و کارایی می‌شوند. برای پرداختن به این مسائل، در مقاله "URetinx-Net: Retinex-based Deep Unfolding Network for Low-light Image Enhancement"^[۲]، نویسندگان یک شبکه بازگشایی عمیق مبتنی بر رتینکس (URetinx-Net) را پیشنهاد می‌کنند که یک مسئله بهینه‌سازی را در یک شبکه قابل یادگیری برای تجزیه تصویر کم‌نور به لایه‌های بازتابی و روشنایی^{۲۴} باز می‌کند. با فرموله کردن مسئله تجزیه به عنوان یک مدل منظم‌سازی شده با پیش‌فرض‌های ضمنی، سه ماژول مبتنی بر یادگیری با دقت طراحی شده‌اند که مسئولیت مقداردهی اولیه وابسته به داده، بهینه‌سازی بازگشایی با کارایی بالا، و افزایش روشنایی مشخص شده توسط کاربر را بر عهده دارند. به ویژه، ماژول بهینه‌سازی بازگشایی پیشنهادی، که دو شبکه را معرفی می‌کند تا به صورت تطبیقی با پیش‌فرض‌های ضمنی به شیوه‌ای مبتنی بر داده‌ها سازگار شوند، می‌تواند سرکوب نویز و حفظ جزئیات را برای نتایج تجزیه نهایی درک کند. آزمایش‌های گسترده بر روی تصاویر کم‌نور دنیای واقعی از نظر کیفی و کمی، اثربخشی و برتری روش پیشنهادی را نسبت به روش‌های پیشرفته نشان می‌دهد.

از آنجایی که نظریه رتینکس به خوبی ادراک رنگی بینایی انسان را مدل می‌کند، روش‌های بهبود تصویر در نور کم^{۲۵} (LLIE) مبتنی بر نظریه رتینکس توجه زیادی را به خود جلب کرده است. همانطور که در نظریه رتینکس بیان شد، یک تصویر را می‌توان به دو جزء، یعنی بازتاب و روشنایی تجزیه کرد. از نظر ریاضی، تصویر مشاهده شده I را می‌توان با فرمول زیر بیان کرد:

$$I = R \cdot L \quad (۷-۲)$$

که در آن R ، L ، و \cdot به ترتیب بازتاب، روشنایی و ضرب عنصری^{۲۶} را نشان می‌دهند. در برخی از روش‌های اولیه مبتنی بر رتینکس، ابتدا روشنایی تخمین زده می‌شود، و سپس بازتاب به عنوان نتایج بهبود یافته

^{۲۳} Retinex model-based methods

^{۲۴} reflectance and illumination

^{۲۵} low-light image enhancement

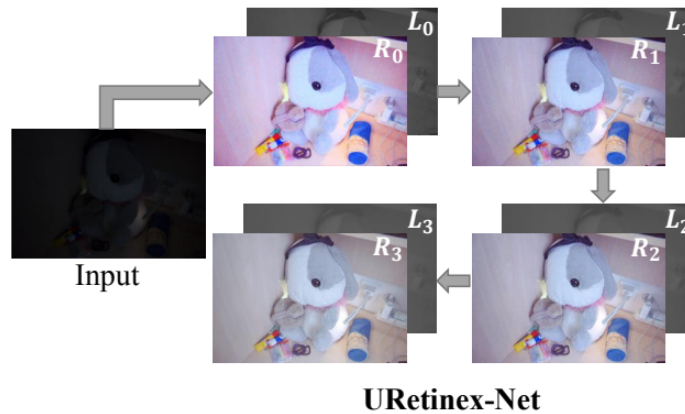
^{۲۶} element-wise multiplication

نهایی در نظر گرفته می‌شود. اگرچه جزئیات را می‌توان تا حد زیادی از ورودی بازبایی کرد، اما اغلب منجر به ظاهری غیر طبیعی و بیش از حد در معرض دید می‌شود. پس از آن، تعدادی از روش‌های مبتنی بر مدل که قابلیت تفسیر خوبی دارند برای حل مشکل تجزیه بد وضع در معادله ۲-۷ پیشنهاد شده‌اند، که در آن پیش‌فرض‌های دست‌ساز مختلف که به عنوان عبارات منظم‌سازی معرفی شده، در مدل‌ها طراحی می‌شوند. سپس، برای نشان دادن تصویر در نور کم، روشنایی با تصحیح گاما^{۲۷} بیشتر می‌شود. طراحی پیش‌فرض ضمنی برای برازش داده‌ها، کلیدی است برای این که مدل‌ها به خوبی کار کنند، اما برای روش‌های مبتنی بر مدل، سازگاری کافی در صحنه‌های مختلف امری چالش برانگیز است. علاوه بر این، اکثر روش‌های مبتنی بر مدل که از طرح‌های بهینه‌ساز تکراری مرسوم استفاده می‌کنند، برای یک تنظیم تصویر منفرد پرهزینه هستند، که مانع توسعه آنها در برنامه‌های بلادرنگ^{۲۸} می‌شود.

با توجه به این محدودیت‌های موجود در روش‌های مبتنی بر مدل، محققان از شبکه‌های عمیق برای بازبایی تصاویر در نور کم به شیوه‌ای مبتنی بر داده استفاده می‌کنند. در میان این روش‌های مبتنی بر یادگیری، روش‌های مبتنی بر رتینکس از شبکه‌های عمیق برای تخمین بازتاب و روشنایی و افزایش روشنایی استفاده می‌کنند. با این حال، اکثر این روش‌ها پس از تجزیه، عملیات حذف نویز را بر روی بازتاب انجام می‌دهند که منجر به از دست رفتن جزئیات می‌شود. علاوه بر این، روش‌های مبتنی بر یادگیری از کمبود تفسیرپذیری و انعطاف‌پذیری رنج می‌برند، که مشکلاتی را در تحلیل محدودیت‌های بالقوه شبکه‌های طراحی‌شده به همراه دارد. برای این منظور، نویسندگان یک شبکه بازگشایی عمیق مبتنی بر رتینکس (URetinex-Net) را پیشنهاد می‌کنند تا تصاویر با نور کم را در فضای رنگی RGB نشان دهد.

برای ادغام نقاط قوت از روش‌های مبتنی بر مدل و مبتنی بر یادگیری، نویسندگان مقاله، مسئله تجزیه مبتنی بر Retinex را به عنوان یک مدل منظم‌سازی پیش‌فرض‌های ضمنی فرمول‌بندی می‌کنند، که در آن عبارات منظم‌سازی قوی به جای استفاده از پیش‌فرض‌های دست‌ساز، توسط شبکه‌های قابل یادگیری استنباط می‌شوند. به طور خاص، تابع انرژی مدل فرموله شده از طریق الگوریتم تقسیم نیم درجه دوم متناوب^{۲۹} به چهار مشکل فرعی تک متغیره تقسیم می‌شود و این مسئله بهینه‌سازی را می‌توان با کمینه‌سازی مکرر چهار زیرمسئله حل کرد. سپس، طرح بهینه‌سازی را در یک شبکه عمیق بازگشایی می‌کنند. برای مسائل فرعی مربوط به عبارات پیش‌فرض‌های قبلی، دو شبکه معرفی می‌شوند تا به

^{۲۷} Gamma correction^{۲۸} real-time^{۲۹} half-quadratic splitting algorithm



شکل ۲-۳: تجزیه منجر به هر مرحله بازگشایی می‌شود، جایی که اجزای تجزیه شده به تدریج از شر تخریب خلاص می‌شوند. اصلاح گاما (Gamma correction) برای لایه روشنایی تجزیه شده برای جلوه بصری بهتر اعمال می‌شود.

صورت تطبیقی با پیش‌فرض‌های ضمنی تناسب داشته باشند، در حالی که بقیه در رابطه با عبارت وفاداری با راه حل‌های نزدیک متناظر حل می‌شوند. در طول بهینه‌سازی بازگشایی، بازتاب تجزیه شده و روشنایی به صورت مرحله‌ای از تخریب اشاره شده خلاص می‌شوند (شکل ۲-۳ را ببینید). در همین حال، مدل فرموله شده از طراحی شرایط پیش‌فرض‌های ضمنی اجتناب می‌کند. علاوه بر این، با توجه به تأثیر مهم مقداردهی اولیه بر بهینه‌سازی، نویسندگان یک مازول مقداردهی اولیه را پیشنهاد می‌کنند که به بهبود بهینه‌ساز کمک کند. در نهایت، یک مازول تنظیم روشنایی را طراحی می‌کنند تا نقشه روشنایی منعطف را با توجه به سطح نور مشخص شده استفاده کند.

۲-۲-۱ الگوریتم متنی بر Retinex برای بهبود کیفیت تصاویر کم نور

روش‌های مبتنی بر مدل: نظریه رتینکس کلاسیک سیستم بینایی انسان^{۳۰} (HVS) را مدل می‌کند، که فرض می‌کند رنگ مشاهده شده به اجزای ذاتی خود جسم بستگی دارد و منبع نور غیریکنواخت بیرونی روی جسم می‌افتد. به طور طبیعی، تصویر را می‌توان به بازتاب و روشنایی تجزیه کرد، همانطور که در معادله ۲-۷ نشان داده شده است. چندین مدل تجزیه رتینکس تحت چارچوب‌های متغیر پیشنهاد شده است. سپس با تنظیم روشنایی تخمین زده شده، تصویر هدف در نور کم بازیابی می‌شود. پس از آن، چندین روش مبتنی بر مدل که توابع انرژی تحت چارچوب حداکثر پسینی^{۳۱} (MAP) قرار دارند، پیشنهاد می‌شوند. به طور کلی، روش‌های مبتنی بر مدل مرسوم، بیشتر بر پیش‌فرض‌های دست‌ساز که

^{۳۰} human visual system

^{۳۱} maximizing a posterior

با دقت طراحی شده یا مدل‌های آماری خاصی تکیه می‌کنند. با این حال، زمانی که در صحنه‌های مختلف اعمال می‌شود، چنین پیش‌فرض‌هایی با ظرفیت مدل محدود می‌شوند.

روش‌های مبتنی بر یادگیری عمیق: در دهه گذشته، روش‌های مبتنی بر یادگیری عمیق نتایج امیدوارکننده‌ای را برای مسائل LLIE ارائه کرده‌اند. با الهام از نظریه رتینکس یک شبکه آموزشی سرتاسری به نام Retinex-Net را پیشنهاد شده که شامل یک مازول تجزیه و یک مازول تنظیم روشنایی است. علاوه بر این، یک تابع نگاشت قابل یادگیری در مازول تنظیم روشنایی طراحی شده است که در آن می‌توان تصاویر را به صورت انعطاف‌پذیر تحت سطح نور خاص کاربر بازیابی کرد. اخیراً با الهام از نظریه رتینکس همراه با حداکثر آنتروپی یک چارچوب خود نظارتی پیشنهاد شده که از عملکرد یکسان‌سازی هیستوگرام برای اعمال محدودیت بر بازتاب استفاده می‌کند. اگرچه این روش‌ها عملکرد قابل توجهی را در LLIE نشان داده‌اند، اما فاقد قابلیت تفسیر هستند که مانع توسعه آنها در LLIE می‌شود. علاوه بر این، بر اساس این تئوری که بازتابی که اجزای ذاتی را به تصویر می‌کشد باید در محیط‌های نوری مختلف سازگار باشد، اکثر روش‌های یادگیری عمیق مبتنی بر رتینکس، بازتاب را پس از تجزیه بازیابی می‌کنند، که منجر به از دست رفتن جزئیات می‌شود.

روش‌های بازگشایی عمیق: روش‌های LLIE مبتنی بر مدل بسیار قابل تفسیر و انعطاف‌پذیر هستند، در حالی که روش‌های LLIE مبتنی بر یادگیری در یادگیری نقشه‌برداری پیچیده به شیوه‌ای مبتنی بر داده برتری دارند. علاوه بر این، شبکه‌های عصبی عمیق در طول استنتاج سریع عمل می‌کنند که به ویژه از نظر محاسباتی کارآمد است. الگوریتم بازگشایی که از نقاط قوت نهفته در روش‌های مبتنی بر مدل و مبتنی بر یادگیری استفاده می‌کند، در دهه گذشته توجه زیادی را به خود جلب کرده است. گرگور و لکون^{۳۲} ابتدا یک الگوریتم بازگشایی زمان را برای حل الگوریتم انقباض و آستانه‌سازی تکراری^{۳۳} در بهینه‌سازی کدگذاری پراکنده^{۳۴} طراحی کردند، به طوری که الگوریتم پیشنهادی عملکرد رقابتی را در تکرارهای کمتری تولید می‌کند. با الهام از چنین طرح بهینه‌سازی، الگوریتم‌های بازگشایی عمیق تأثیر زیادی بر بسیاری از مسائل مهم پردازش تصویر، مانند وضوح فوق‌العاده، حذف نویز تصویر، سرکوب به هم ریختگی و حذف نویز گذاشته‌اند. اخیراً یک چارچوب آشکار را برای تخمین روشنایی و حذف نویز به روشی بدون نظارت پیشنهاد کردند، در حالی که ارتباط متقابل بین بازتاب و روشنایی در این راه نادیده گرفته خواهد شد. روش ما از دو جنبه اصلی با آن متفاوت است: (۱) ما تمایل داریم به طور

Gregor and Lecun^{۳۲}iterative shrinkage and thresholding^{۳۳}sparse coding^{۳۴}

همزمان بازتاب و روشنایی ورودی را در یک چارچوب یکپارچه تخمین بزنیم. (۲) شبکه ما می‌تواند به طور انعطاف پذیری روشنایی را از طریق یک نسبت تعریف شده توسط کاربر افزایش دهد.

۲-۲-۲ معماری شبکه URetinex-Net

مدل کلاسیک مبتنی بر رتینکس فرض می‌کند که تصویر را می‌توان از طریق معادله ۷-۲ به بازتاب و روشنایی تجزیه کرد، و پیش‌فرض‌های دست‌ساز مختلفی برای حل این مشکل تجزیه نامطلوب تحت چارچوب MAP ایجاد شده‌اند. بنابراین، بازتاب و روشنایی را می‌توان با به حداقل رساندن تابع انرژی منظم زیر به دست آورد:

$$E(R, L) = \|I - R \cdot L\|_F^2 + \alpha\Phi(R) + \beta\Psi(L) \quad (۸-۲)$$

که در آن $\|\cdot\|_F^2$ نشان دهنده نرم فربنیوس^{۳۵} است، $\|I - R \cdot L\|_F^2$ عبارت وفاداری است که از معادله ۷-۲ مشتق شده است. $\alpha\Phi(R)$ و $\beta\Psi(L)$ عبارات منظم‌سازی هستند که به ترتیب پیش‌فرض‌های تحمیلی بر R و L را نشان می‌دهند و α و β پارامترهای سبک-سنگینی^{۳۶} هستند. به طور کلی، برای تسهیل بهینه‌سازی، عبارات وفاداری و عبارات منظم‌سازی به طور جداگانه مورد بررسی قرار می‌گیرند، به طوری که ما دو متغیر کمکی P و Q را به ترتیب برای تقریب R و L معرفی می‌کنیم. بر این اساس، این منجر به مشکل کمینه سازی زیر می‌شود:

$$\begin{aligned} \min_{P, Q, R, L} & \|I - P \cdot Q\|_F^2 + \alpha\Phi(R) + \beta\Psi(L) \\ \text{s.t.} & P = R, Q = L \end{aligned} \quad (۹-۲)$$

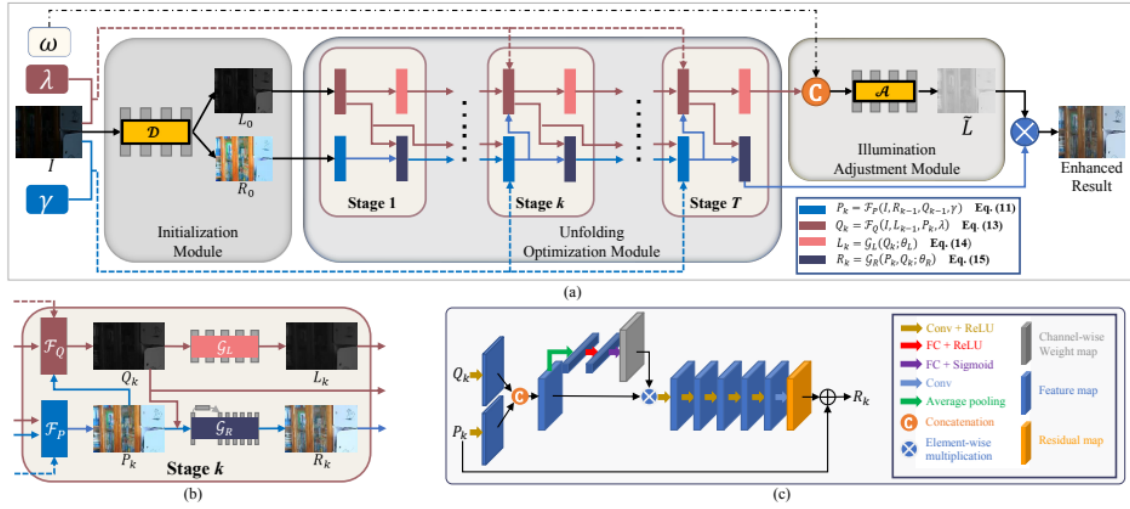
برای مقابله با محدودیت‌های برابری، دو عبارت جریمه درجه دوم معرفی می‌شوند و مسئله به صورت بازنویسی می‌شود:

$$\min_{P, Q, R, L} \|I - P \cdot Q\|_F^2 + \alpha\Phi(R) + \beta\Psi(L) + \gamma\|P - R\|_F^2 + \lambda\|Q - L\|_F^2 \quad (۱۰-۲)$$

که در آن γ و λ پارامترهای جریمه هستند.

^{۳۵} Frobenius

^{۳۶} trade-off



شکل ۲-۴: تصویر URetinex-Net پیشنهادی. (الف) چارچوب کلی URetinex-Net، (ب) جزئیات در مورد هر مرحله در URetinex-Net، (ج) ساختار شبکه خاص \mathcal{G}_R اعمال شده در هر مرحله. به ویژه، برای مشکل LLIE شامل سه مازول قابل یادگیری است. با ارسال یک تصویر هدف در نور کم به مازول اولیه، بازتاب اولیه و روشنایی ایجاد می‌شود. پس از آن، مازول بهینه‌سازی باز شده لایه‌های بازتاب و روشنایی را به طور مکرر اصلاح می‌کند. در نهایت، مازول تنظیم روشنایی با توجه به نسبت تعریف شده توسط کاربر، نسخه بهبودیافته با نور معمولی را تولید می‌کند.

برای حل مسئله در معادله ۲-۱۰، مقدار یک متغیر به طور متناوب با متغیرهای ثابت بروز می‌شود. بنابراین، ما مسئله را به چهار زیرمسئله تک متغیره تقسیم می‌کنیم که می‌تواند با طرح متناوب زیر بهینه شود که در آن k نشان دهنده شاخص تکرار است:

$$P_k = \arg \min_P \|I - P \cdot Q_{k-1}\|_F^2 + \gamma \|P - R_{k-1}\|_F^2 \quad (2-11)$$

$$R_k = \arg \min_R \alpha \Phi(R) + \gamma \|P_k - R\|_F^2 \quad (2-12)$$

$$Q_k = \arg \min_Q \|I - P_k \cdot Q\|_F^2 + \lambda \|Q - L_{k-1}\|_F^2 \quad (2-13)$$

$$L_k = \arg \min_L \beta \Psi(L) + \lambda \|Q_k - L\|_F^2 \quad (2-14)$$

از آنجایی که طراحی عبارات منظم‌سازی خاص $\alpha \Phi(R)$ و $\beta \Psi(L)$ دشوار است، ما از شبکه‌های عمیق برای تطبیق با اولویت‌های فیزیکی R و L استفاده می‌کنیم. بنابراین، بر اساس طرح بهینه‌سازی ذکر شده در بالا، مراحل بروزرسانی را ترسیم می‌کنیم. به یک معماری شبکه بازگشایی عمیق، و یک

چارچوب جدید برای LLIE پیشنهاد کنید. همانطور که در شکل ۲-۴ نشان داده شده است، URetinex-Net پیشنهادی شامل سه ماژول، یعنی ماژول مقداردهی اولیه، ماژول بهینه‌سازی بازگشایی، و ماژول تنظیم روشنایی است.

۳-۲ ترمیم تصاویر (Image restoration)

در مقاله "Deep Generalized Unfolding Networks for Image Restoration" [۳]، نویسندگان یک شبکه بازگشایی تعمیم‌یافته عمیق (DGUNet) برای بازیابی تصویری پیشنهاد می‌کنند. به طور مشخص، بدون از دست دادن قابلیت تفسیر، آنها یک استراتژی تخمین گرادیان را در مرحله نزول گرادیان الگوریتم گرادیان کاهشی نزدیک^{۳۷} (PGD) ادغام می‌کنند، که آن را به سمت مقابله با تخریب تصویر پیچیده و واقعی سوق می‌دهد. علاوه بر این، نویسندگان مسیرهای اطلاعاتی بین مرحله‌ای را در سراسر نقشه‌برداری نزدیک در تکرارهای مختلف PGD طراحی می‌کنند تا از دست رفتن اطلاعات ذاتی در اکثر شبکه‌های بازگشایی عمیق از طریق یک روش چند مقیاسی^{۳۸} و فضایی-تطبیقی^{۳۹} اصلاح شود. با ادغام گرادیان کاهشی انعطاف‌پذیر و نقشه‌برداری نزدیک اطلاعاتی، الگوریتم PGD تکراری را در یک DNN قابل آموزش بازگشایی می‌کنند. آزمایش‌های گسترده بر روی وظایف مختلف بازیابی تصویر، برتری این روش را از نظر عملکرد، تفسیرپذیری و تعمیم‌پذیری، نشان می‌دهد.

روش‌های مبتنی بر مدل در بازیابی تصویر (IR) معمولاً آنرا به عنوان یک مسئله بیزی^{۴۰} فرموله کرده و معادله را حل می‌کنند. تحت یک چارچوب یکپارچه MAP (به حداکثر رساندن یک پسین^{۴۱}):

$$\hat{x} = \underset{x}{\operatorname{argmax}} \log P(x|y) = \underset{x}{\operatorname{argmax}} \log P(y|x) + \log P(x) \quad (۱۵-۲)$$

که در آن $\log P(y|x)$ و $\log P(x)$ به ترتیب عبارات وفاداری داده‌ها و عبارت منظم‌سازی را نشان می‌دهد. عبارت وفاداری داده معمولاً به عنوان یک نرم ℓ_2 تعریف می‌شود و معادله ۲-۱۵ را به عنوان تابع انرژی

^{۳۷} Proximal Gradient Descent

^{۳۸} multi-scale

^{۳۹} spatial-adaptive

^{۴۰} Bayesian

^{۴۱} maximizing a posterior

زیر بیان می‌کند:

$$x = \underset{x}{\operatorname{argmin}} \frac{1}{2} \|y - Ax\|_2^2 + \lambda J(x) \quad (۱۶-۲)$$

که در آن λ یک فرایارامتر برای وزن دادن به عبارت منظم‌سازی $J(x)$ است. عبارت وفاداری داده تضمین می‌کند که راه حل با تخریب مطابقت دارد. عبارت منظم‌سازی، مسئله بد وضع معکوس را با اعمال ویژگی مورد نظر که شامل پیش‌فرض‌های پیچیده است، برطرف می‌کند. با این حال، توانایی نمایش طراحی دست‌ساز محدود است که منجر به نتایج ناپایدار می‌شود و معمولاً در استنتاج زمان‌بر هستند.

اخيراً بازیابی تصویر (IR) با یادگیری عمیق به موفقیت چشمگیری دست یافته است، زیرا آنها می‌توانند از مجموعه داده‌های در مقیاس بزرگ، پیش‌فرض‌های قوی را بیاموزند. با وجود عملکرد امیدوارکننده و استنتاج سریع، طراحی جعبه سیاه گونه آن‌ها تجزیه و تحلیل نقش اجزای مختلف را دشوار می‌کند و دستاوردهای عملکردی آن‌ها اغلب به چیدن مازول‌های جدید (به قیمت افزایش پیچیدگی مدل) نسبت داده می‌شود.

برای ترکیب تفسیرپذیری و سازگاری، برخی از روش‌های ترکیبی^{۴۲}، ادغام شبکه‌های عمیق را در الگوریتم‌های بهینه‌سازی کلاسیک پیشنهاد کردند. به عنوان مثال، روش‌های عمیق plug-and-play (PNP)، حذف‌کننده نویز CNN از قبل آموزش‌دیده را به عنوان پیش‌فرض قبلی، در چارچوب‌های بهینه‌سازی تکراری برای وظایف مختلف IR یکپارچه کردند. متأسفانه آنها معمولاً از استنباط وقت گیر رنج می‌برند.

اخيراً، شبکه‌های بازگشایی عمیق (DUN) بهینه‌سازی تمام پارامترها را به صورت سرتاسری^{۴۳}، ارائه عملکرد بهتر و استنتاج سریع‌تر پیشنهاد کردند. با این حال، به دلیل طراحی قابل تفسیر، اکثر آنها به فرآیندهای تخریب شناخته شده برای استخراج راه حل نیاز دارند. با این وجود، فرآیندهای تخریب در برنامه‌های کاربردی دنیای واقعی با توزیع متغیر وابسته به سیگنال و فضایی پیچیده و نامطمئن هستند. بنابراین، اکثر روش‌های DUN مفروضات تخریب دست ساز را ایجاد می‌کنند یا به صراحت کار خالص را با عوامل تخریب برای مقابله با مشکلات تخریب تصویر از پیش تعریف شده ارائه می‌کنند. علاوه بر این، از آنجایی که روش‌های سنتی مبتنی بر مدل یک تصویر را در هر تکرار خروجی می‌دهند، DUN مربوطه باید ورودی و خروجی هر مرحله را به عنوان یک تصویر بپذیرد. این طراحی ذاتی به ناچار منجر

^{۴۲} hybrid methods

^{۴۳} end-to-end

به تحریف اطلاعات ویژگی به تصویر می‌شود. به چنین از دست رفتن اطلاعات در DUN در آثار موجود توجه کمی شده است.

برای اصلاح مسائل فوق و پر کردن شکاف بین روش‌های مبتنی بر مدل و روش‌های یادگیری عمیق، ما یک شبکه بازگشایی تعمیم‌یافته عمیق (DGUNet) را پیشنهاد می‌کنیم. از یک طرف، روش ما به عنوان روش‌های مبتنی بر مدل با فرمول‌بندی طراحی مدل از طریق الگوریتم گرادیان کاهشی نزدیک (PGD) قابلیت تفسیر خوبی دارد. از سوی دیگر، مشابه روش‌های یادگیری عمیق، روش ما به صورت سرتاسری با یک مسیر ویژگی بدون مانع آموزش داده می‌شود و به راحتی می‌توان آن را برای برنامه‌های پیچیده و دنیای واقعی اعمال کرد. برای دستیابی به این هدف، ابتدا یک استراتژی تخمین گرادیان را با مرحله گرادیان کاهشی الگوریتم PGD ادغام می‌کنیم تا گرادیان را در موارد ناشناخته تخریب پیش‌بینی کنیم. سپس مسیرهای اطلاعاتی بین مرحله‌ای را برای جبران از دست دادن اطلاعات ذاتی در DUN طراحی می‌کنیم.

۱-۳-۲ الگوریتم سنتی Proximal Gradient Descent

از نظر فنی، الگوریتم PGD تقریباً معادله ۱۶-۲ را به عنوان یک مسئله همگرایی تکراری از طریق تابع تکراری زیر بیان می‌کند:

$$\hat{\mathbf{x}}^k = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - (\hat{\mathbf{x}}^{k-1} - \rho \nabla g(\hat{\mathbf{x}}^{k-1}))\|_2^2 + \lambda J(\mathbf{x}) \quad (17-2)$$

که در آن $\hat{\mathbf{x}}_k$ به خروجی تکرار k -ام اشاره دارد، و $g(\cdot)$ عبارت وفاداری داده را در معادله ۱۶-۲ نشان می‌دهد. ∇ عملگر دیفرانسیل است که با اندازه گام ρ وزن می‌شود.

از نظر ریاضی، قسمت $(\hat{\mathbf{x}}^{k-1} - \rho \nabla g(\hat{\mathbf{x}}^{k-1}))$ تابع فوق یک عملیات گرادیان کاهشی است و قسمت بقیه فرمول را می‌توان با عملگر نزدیکی $\text{prox}_{\lambda, J}$ حل کرد. بنابراین، منجر به دو مشکل فرعی می‌شود، یعنی گرادیان کاهشی (معادله ۱۸-۲) و نقشه برداری نزدیک (معادله ۱۹-۲):

$$\mathbf{v}^k = \mathbf{x}^{k-1} - \rho A^T (A \hat{\mathbf{x}}^{k-1} - \mathbf{y}), \quad (18-2)$$

$$\hat{\mathbf{x}}^k = \text{prox}_{\lambda, J}(\mathbf{v}^k). \quad (19-2)$$

الگوریتم PGD به طور مکرر \mathbf{v}^k و $\hat{\mathbf{x}}^k$ را تا زمان همگرایی بروز می‌کند. ISTA یک الگوریتم معمولی مبتنی بر PGD است که در آن عبارت منظم‌ساز به عنوان یک نرم ℓ_1 تعریف می‌شود، یعنی $J(x) = \|x\|_1$. بنابراین، نگاشت نزدیک در ISTA به عنوان یک تابع آستانه نرم مشتق شده است:

$$\text{prox}_{\lambda, J}(\mathbf{v}^k) = \text{sign}(\mathbf{v}^k) \max(0, |\mathbf{v}^k| - \lambda)$$

با این حال، نرم ℓ_1 دست ساز توانایی‌های محدودی برای نمایش دارد و کاربرد آن به چند کار شناخته شده تخریب محدود می‌شود. با تمرکز بر بهبود الگوریتم سنتی PGD، در این مقاله، آن را توسط شبکه‌های عصبی عمیق با طراحی قوی و تعمیم یافته بازگشایی می‌کنیم.

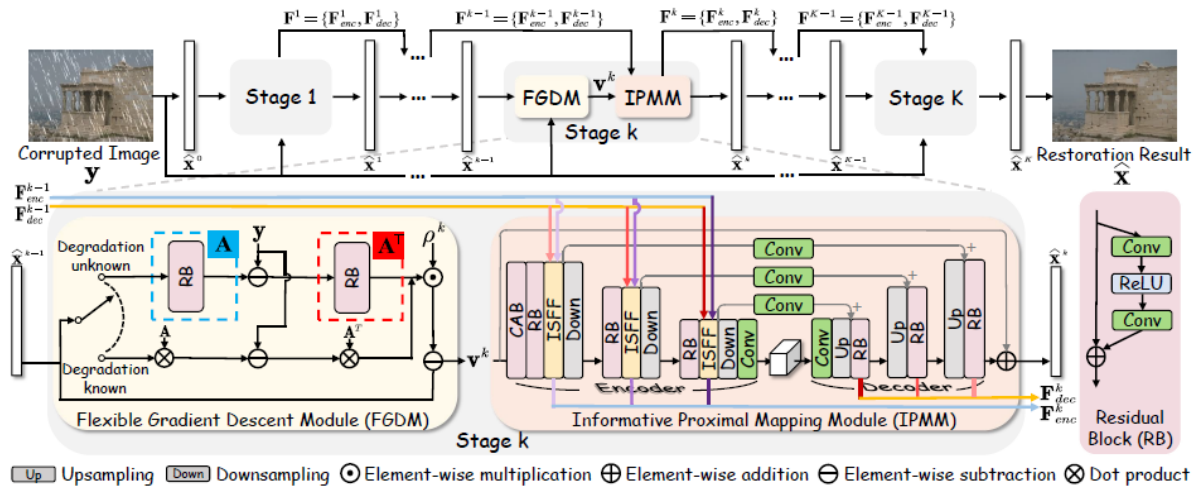
۲-۳-۲ معماری شبکه DGUNet

کل معماری شبکه DGUNet پیشنهادی ما در شکل ۲-۵ ارائه شده است که یک چارچوب بازگشایی از الگوریتم گرادیان نزولی نزدیک (PGD) بر اساس شبکه‌های عصبی عمیق (DNN) است. DGUNet ما از چندین مرحله تکراری تشکیل شده است. هر مرحله شامل یک ماژول گرادیان کاهشی منعطف^{۴۴} (FGDM) و یک ماژول نگاشت نزدیک اطلاعاتی^{۴۵} (IPMM) است که به ترتیب مربوط به گرادیان کاهشی (معادله ۲-۱۸) و نقشه برداری نزدیک (معادله ۲-۱۹) در یک مرحله تکرار الگوریتم PGD است. تعداد مراحل به طور پیش فرض هفت تنظیم شده است و به جز مرحله اول و آخر، پارامترهای یکسانی را به اشتراک می‌گذارند. برای بهبود بیشتر عملکرد مدل، نویسندگان همچنین یک نسخه پلاس با نام DGUNet+ ارائه می‌دهند که در آن تمام مراحل مستقل از پارامتر هستند.

ماژول گرادیان کاهشی انعطاف پذیر: همانطور که در معادله ۲-۱۸ نشان داده شده است، مرحله گرادیان کاهشی زمانی که ماتریس تخریب A شناخته شده باشد، بی‌اهمیت است. با این حال، A در برخی از مسائل تخریب ناشناخته است، و محاسبه گرادیان (یعنی $A^T(\hat{\mathbf{x}}^{k-1} - \mathbf{y})$) غیرقابل حل است. در این زمینه، ما یک ماژول گرادیان کاهشی انعطاف پذیر (FGDM) را پیشنهاد می‌کنیم که در ردیف دوم شکل ۲-۵ نشان داده شده است. این ماژول دارای دو تنظیمات مدل برای مقابله با تخریب موارد شناخته شده و ناشناخته است. در موردی که A شناخته شده است، ما مستقیماً از A دقیق برای محاسبه گرادیان استفاده می‌کنیم. برای بهبود ثبات، اندازه گام ρ را به عنوان یک پارامتر قابل آموزش در هر

^{۴۴} flexible gradient descent module

^{۴۵} formative proximal mapping module



شکل ۲-۵: تصویری از شبکه آشکارسازی تعمیم یافته عمیق پیشنهادی ما (DGUNet). ما معماری کلی را در ردیف اول ارائه می‌کنیم که عمدتاً از چندین مرحله تشکیل شده است. هر مرحله مربوط به یک تکرار در الگوریتم PGD است. ردیف دوم طراحی دقیق هر مرحله را ارائه می‌کند که شامل یک ماژول گرادیان کاهشی انعطاف‌پذیر (FGDM) و یک ماژول نگاشت نزدیک اطلاعاتی (IPMM) است.

مرحله تنظیم می‌کنیم که منجر به عملیات گرادیان کاهشی زیر می‌شود:

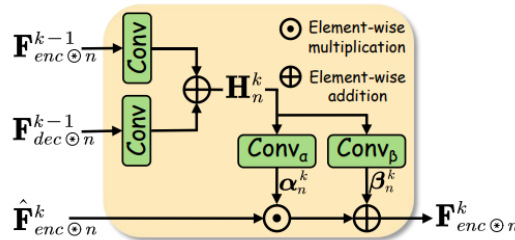
$$\mathbf{v}^k = \hat{\mathbf{x}}^{k-1} - \rho^k \mathbf{A}^\top (\mathbf{A} \hat{\mathbf{x}}^{k-1} - \mathbf{y}) \quad (20-2)$$

اگر \mathbf{A} ناشناخته باشد، به جای ایجاد فرضیات خاص برای مشکلات مختلف تخریب، یک استراتژی مبتنی بر داده را برای پیش‌بینی گرادیان اتخاذ می‌کنیم. از نظر فنی، ما از دو بلوک باقیمانده مستقل، با نام‌های $\mathcal{F}_{\mathbf{A}}^k$ و $\mathcal{F}_{\mathbf{A}^\top}^k$ برای شبیه‌سازی \mathbf{A} و ترنسپوز آن \mathbf{A}^\top در مرحله k -ام استفاده می‌کنیم. گرادیان به صورت $\mathcal{F}_{\mathbf{A}^\top}^k (\mathcal{F}_{\mathbf{A}}^k (\hat{\mathbf{x}}^{k-1}) - \mathbf{y})$ محاسبه می‌شود. بنابراین بدون از دست دادن قابلیت تفسیر، گرادیان کاهشی در DGUNet پیشنهادی ما می‌تواند به عنوان تابع زیر در موارد تخریب ناشناخته تعریف شود.

$$\mathbf{v}^k = \hat{\mathbf{x}}^{k-1} - \rho^k \mathcal{F}_{\mathbf{A}^\top}^k (\mathcal{F}_{\mathbf{A}}^k (\hat{\mathbf{x}}^{k-1}) - \mathbf{y}) \quad (21-2)$$

ماژول نقشه برداری نزدیک اطلاعاتی: برای حل معادله ۲-۱۹، مشخص شده است که از دیدگاه بیزی، در واقع با یک مسئله حذف نویز مطابقت دارد. در این زمینه، ما یک ماژول نقشه برداری نزدیک اطلاعاتی (IPMM) را طراحی می‌کنیم که در ردیف دوم شکل ۲-۵ نشان داده شده. IPMM ما یک

معماری ساعت شنی شکل است که از یک رمزگذار^{۴۶} و یک رمزگشا^{۴۷} برای استفاده از نقشه‌های ویژگی چند مقیاسی^{۴۸} تشکیل شده است.



شکل ۲-۶: تصویری از ماژول همجوشی ویژگی بین مرحله‌ای ما (ISFF) در مقیاس n -ام در مرحله k -ام. ویژگی‌های رمزگذار و رمزگشا از مرحله قبل به صورت عادی سازی فضایی-تطبیقی با مرحله فعلی ترکیب می‌شوند.

به طور خاص، IPMM ما با یک بلوک توجه کانال^{۴۹} (CAB) برای استخراج ویژگی‌های کم عمق شروع می‌شود. ما از بلوک باقیمانده^{۵۰} (RB) بدون نرمال سازی دسته‌ای^{۵۱} برای استخراج ویژگی‌ها در سه مقیاس استفاده می‌کنیم. در اینجا، ما از $\mathbf{F}_{enc}^k = \{\mathbf{F}_{enc \odot n}^k\}_{n=1}^3$ و $\mathbf{F}_{dec}^k = \{\mathbf{F}_{dec \odot n}^k\}_{n=1}^3$ برای نمایش ویژگی‌های رمزگذار و رمزگشا استخراج شده از مرحله k -ام در مقیاس n -ام. به منظور تغییر مقیاس در IPMM خود، ما از max-pooling با ابعاد 2×2 با stride ۲ برای نمونه برداری با وضوح پایین^{۵۲} استفاده می‌کنیم، و از نمونه برداری با وضوح بالا^{۵۳} دوخطی و به دنبال آن یک لایه کانولوشن برای نمونه برداری استفاده می‌کنیم. مشابه بسیاری از حذف‌کننده‌های نويز رقابتی، یک مسیر سراسری از ورودی به خروجی اضافه می‌کنیم که شبکه را تشویق می‌کند تا اطلاعات با تکرار^{۵۴} پایین را دور بزند. در پایان IPMM، ما از ماژول توجه با ناظر^{۵۵} (SAM) برای استخراج ویژگی‌های تمیز و سپس تزریق آنها به مرحله بعدی از طریق طرح‌ریزی زیرفضا^{۵۶} استفاده می‌کنیم.

با توجه به از دست دادن اطلاعات ذاتی در اکثر روش‌های DUN، ما مسیرهای اطلاعاتی بین مرحله‌ای را در هر مقیاس طراحی می‌کنیم تا اطلاعات مفیدی را از رمزگذار و رمزگشا در مراحل مختلف پخش

- encoder^{۴۶}
- decoder^{۴۷}
- multi-scale feature map^{۴۸}
- channel attention block^{۴۹}
- residual block^{۵۰}
- batch normalization^{۵۱}
- downsampling^{۵۲}
- upsampling^{۵۳}
- frequency^{۵۴}
- supervised attention module^{۵۵}
- subspace projection^{۵۶}

کنیم. برای هدف تصویر، از خطوط رنگی مختلف برای تشخیص اطلاعات رمزگذار و رمزگشا با مقیاس‌های مختلف در شکل ۵-۲ استفاده می‌کنیم. برای ترکیب اطلاعات بین مرحله‌ای، ما یک زیر ماژول همجوشی ویژگی بین مرحله‌ای^{۵۷} (ISFF) در هر مقیاس در رمزگذار طراحی می‌کنیم. توجه داشته باشید که به دلیل پرش اتصالات بین رمزگذار و رمزگشا، اطلاعات بین مرحله‌ای نیز می‌تواند به طور طبیعی به رمزگشا منتقل شود. معماری دقیق ISFF ما در شکل ۶-۲ ارائه شده است. بطور مشخص، در هر مقیاس، ویژگی‌های رمزگذار و رمزگشا را از مرحله قبل به مرحله فعلی منتقل می‌شود. آنها ابتدا توسط دو لایه کانولوشنال 1×1 مستقل جاسازی شده و با جمع عنصری ادغام می‌شوند. در مرحله k -ام، نتیجه همجوشی در مقیاس n به صورت \mathbf{H}_n^k نشان داده می‌شود. برای محاسبه دو پارامتر وابسته $\alpha_n^k, \beta_n^k \in \mathbb{R}^{C \times H \times W}$ برای انتقال خروجی میانی $\hat{\mathbf{F}}_{enc \otimes n}^{k-1} \in \mathbb{R}^{C \times H \times W}$ به یک پارامتر اطلاعاتی $\mathbf{F}_{enc \otimes n}^{k-1} \in \mathbb{R}^{C \times H \times W}$ استفاده می‌شود، که در آن C, H, W به ترتیب به اندازه کانال، ارتفاع و عرض اشاره دارد. از نظر ریاضی، ترکیب ویژگی‌های بین مرحله‌ای پیشنهادی ما به عنوان نمایش ویژگی زیر تعریف می‌شود:

$$\begin{cases} \mathbf{H}_n^{k-1} &= Conv(\mathbf{F}_{enc \otimes n}^{k-1}) + Conv(\mathbf{F}_{dec \otimes n}^{k-1}) \\ \alpha_n^k, \beta_n^k &= Conv_\alpha(\mathbf{H}_n^{k-1}), Conv_\beta(\mathbf{H}_n^{k-1}) \\ \mathbf{F}_{enc \otimes n}^k &= \hat{\mathbf{F}}_{enc \otimes n}^{k-1} \odot \alpha_a^k + \beta_a^k \end{cases} \quad (22-2)$$

فرآیند همجوشی ویژگی فوق یک نرمال سازی فضایی-تطبیقی استاندارد است. برخلاف روش‌های عادی سازی شرطی، α_n^k و β_n^k بردار نیستند بلکه تانسورهای با ابعاد فضایی هستند. به این ترتیب، در حالی که رمزگذار و رمزگشا ویژگی‌های چند مقیاسی را به دست می‌آورند، نقشه‌های ویژگی در هر مقیاس نیز می‌توانند حافظه تصفیه‌شده مراحل قبلی را با اطلاعات فضایی به خوبی حفظ شده داشته باشند، که منجر به نگاشت نزدیک آموزنده می‌شود. برای هدف تصویرسازی، ما از \mathbf{F}^k برای نمایش مجموعه‌ای از ویژگی‌های رمزگذار و رمزگشای چند مقیاسی استفاده می‌کنیم، به عنوان مثال، $\mathbf{F}^k = \{\mathbf{F}_{enc}^k, \mathbf{F}_{dec}^k\}$. در نهایت، IPMM ما معادله ۱۹-۲ را به صورت زیر بیان می‌کند:

$$\hat{\mathbf{x}}^k, \mathbf{F}^k = \text{prox}_{\theta^k}(\mathbf{v}^k, \mathbf{F}^{k-1}) \quad (23-2)$$

که در آن θ^k پارامترهای IPMM را در مرحله k نشان می‌دهد. در پرتو بحث فوق، در نهایت فرآیند

^{۵۷} inter-stage feature fusion

همگرایی DGUNet خود را در الگوریتم ۱ تعریف می‌کنیم. (شکل ۷-۲ را ببینید.)

Algorithm 1: Proposed DGUNet

Initialization:

- (1) Initialize the iteration depth $k=0$ and ceiling K ;
- (2) Initialize the input $\hat{x}^0 = y$;
- (3) Initialize the inter-stage feature $F^0 = None$;

while $k < K$ **do** **if** A *is unknown* **then** | Update v^{k+1} by Eq. (6); **else** | Update v^{k+1} by Eq. (7); **end** Update \hat{x}^{k+1} and F^{k+1} by Eq. (9); $k = k + 1$;**end****Output:** $[\hat{x}^1, \hat{x}^2, \dots, \hat{x}^K]$

شکل ۷-۲

۴-۲ افزایش وضوح تصاویر (Image super-resolution)

روش‌های مبتنی بر آموزش وضوح تک تصویر^{۵۸} (SISR) به طور مداوم اثربخشی و کارایی برتر را نسبت به روش‌های مبتنی بر مدل سنتی نشان می‌دهند، که عمدتاً به دلیل آموزش سرتاسری است. با این حال، متفاوت از روش‌های مبتنی بر مدل که می‌توانند مشکل SISR را با عوامل مقیاس مختلف، هسته‌های تاری و سطوح نویز تحت یک چارچوب یکپارچه به حداکثر رساندن یک پسین^{۵۹} (MAP) مدیریت کنند، روش‌های مبتنی بر یادگیری عموماً فاقد چنین انعطاف‌پذیری هستند. برای پرداختن به این موضوع، مقاله "Deep Unfolding Network for Image Super-Resolution" [۴] یک شبکه بازگشایی آموزش‌پذیر سرتاسری را پیشنهاد می‌کند که از روش‌های مبتنی بر یادگیری و روش‌های مبتنی بر مدل استفاده می‌کند. به طور خاص، با آشکار کردن استنتاج MAP از طریق یک الگوریتم تقسیم نیمه درجه دوم، تعداد ثابتی از تکرارهای متشکل از حل متناوب یک زیرمسئله داده و یک مشکل فرعی قبلی را می‌توان به دست آورد. سپس دو مشکل فرعی را می‌توان با مازول‌های عصبی حل کرد، که منجر به ایجاد یک شبکه تکراری و آموزش‌پذیر سرتاسری می‌شود. در نتیجه، شبکه پیشنهادی انعطاف‌پذیری روش‌های مبتنی بر

^{۵۸} single image super-resolution^{۵۹} maximizing a posterior

مدل را به ارث می‌برد تا تصاویر تار و دارای نویز را برای فاکتورهای مقیاس مختلف از طریق یک مدل واحد حل کند، در حالی که مزایای روش‌های مبتنی بر یادگیری را حفظ می‌کند. آزمایش‌های گسترده برتری شبکه بازگشایی عمیق پیشنهادی را از نظر انعطاف‌پذیری، اثربخشی و همچنین تعمیم‌پذیری نشان می‌دهد.

افزایش وضوح فوق‌العاده تک تصویر^{۶۰} (SISR) به فرآیند بازیابی نمونه طبیعی و واضح با وضوح بالا^{۶۱} (HR) از یک تصویر با وضوح پایین^{۶۲} (LR) اشاره دارد. این، یکی از مسائل معکوس کلاسیک در بینایی کامپیوتری سطح پایین است و دارای طیف گسترده‌ای از برنامه‌های کاربردی در دنیای واقعی است، مانند افزایش کیفیت بصری تصویر در نمایشگرهای با وضوح بالا و بهبود عملکرد سایر وظایف بینایی سطح بالا. علیرغم دهه‌ها مطالعه، SISR همچنان به مطالعات بیشتر برای اهداف دانشگاهی و صنعتی نیاز دارد. مشکل، عمدتاً ناشی از ناسازگاری بین فرض تخریب ساده روش‌های SISR موجود و تخریب‌های پیچیده تصاویر واقعی است. در واقع، برای ضریب مقیاس s ، مدل تخریب کلاسیک (سنتی) SISR فرض می‌کند که تصویر LR y یک نسخه تار، تخریب‌شده و نویزدار از یک تصویر HR x است. از نظر ریاضی می‌توان آن را با

$$y = (x \otimes k) \downarrow_s + n \quad (24-2)$$

که در آن \otimes نشان‌دهنده پیچش دو بعدی^{۶۳} x با کرنل تار k است، \downarrow_s نشان دهنده نمونه برداری با وضوح پایین استاندارد s -fold است، به عنوان مثال، پیکسل بالا سمت چپ را برای هر پیک $s \times s$ متمایز نگه می‌دارد و بقیه را کنار می‌گذارد، و n معمولاً به عنوان نویز افزودنی گاوسی سفید^{۶۴} (AWGN) که با انحراف معیار استاندارد (یا سطح نویز) σ مشخص شده است، در نظر گرفته می‌شود. با یک معنای فیزیکی واضح، معادله ۲۴-۲ می‌تواند انواع تصاویر LR را با تنظیم هسته‌های تار^{۶۵} مناسب، فاکتورهای مقیاس^{۶۶} و نویزها برای تصاویر متناظر HR تقریب بزند.

به طور خاص، معادله ۲۴-۲ به طور گسترده در روش‌های مبتنی بر مدل مورد مطالعه قرار گرفته

^{۶۰} single image super-resolution

^{۶۱} High-resolution

^{۶۲} Low-resolution

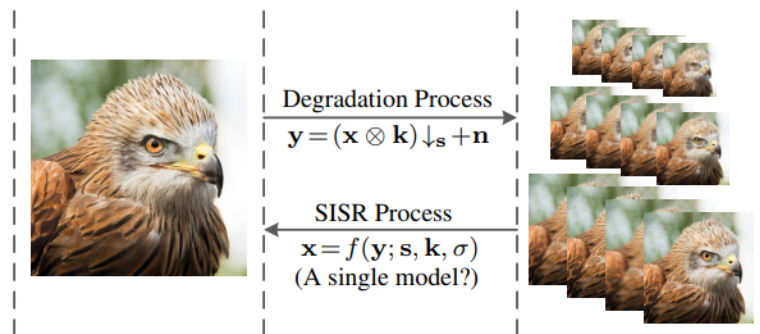
^{۶۳} two-dimensional convolution

^{۶۴} additive, white Gaussian noise

^{۶۵} blur kernels

^{۶۶} scale factors

است که ترکیبی از یک عبارت داده و یک عبارت پیش‌فرض قبلی را تحت چارچوب MAP حل می‌کند. اگرچه روش‌های مبتنی بر مدل معمولاً از نظر الگوریتمی قابل تفسیر هستند، اما معمولاً فاقد معیار استاندارد برای ارزیابی خود هستند، زیرا، جدای از ضریب مقیاس، معادله ۲-۲۴ علاوه بر این شامل یک هسته تاری و نویز اضافه می‌شود. برای راحتی، محققان بدون در نظر گرفتن هسته تاری و سطح نویز به تجزیه دو مکعبی متوسل می‌شوند. با این حال، تخریب دو مکعبی از نظر ریاضی پیچیده است، که به نوبه خود مانع توسعه روش‌های مبتنی بر مدل می‌شود.



شکل ۲-۸: در حالی که یک مدل تخریب منفرد (یعنی معادله ۲-۲۴) می‌تواند منجر به تصاویر LR مختلف برای یک تصویر HR با هسته‌های تاری متفاوت، فاکتورهای مقیاس و نویز، شود، مطالعات بر روی یادگیری یک مدل عمیق واحد برای معکوس کردن همه این تصاویر LR به تصویر HR هنوز کم است.

روش‌های مبتنی بر شبکه عصبی پیچشی (CNN) به دلیل ظرفیت یادگیری بالا و سرعت محاسبات موازی از محبوبیت بالایی برخوردار هستند. با این وجود، کار کمی بر روی استفاده از CNN برای حل معادله ۲-۲۴ از طریق یک مدل واحد انجام شده است. برخلاف روش‌های مبتنی بر مدل، CNN‌ها معمولاً فاقد انعطاف‌پذیری برای افزایش وضوح فوق‌العاده تصاویر LR تار و نویز دار برای فاکتورهای مقیاس مختلف از طریق یک مدل آموزش‌دیده انتها به انتها هستند (شکل ۲-۵ را ببینید).

در این مقاله، نویسندگان یک شبکه با وضوح فوق‌العاده بازگشایی شده (USRNet) را پیشنهاد می‌کنند تا شکاف بین روش‌های مبتنی بر یادگیری و روش‌های مبتنی بر مدل را پر کند. از یک طرف، مشابه روش‌های مبتنی بر مدل، USRNet می‌تواند به طور موثر مدل تخریب کلاسیک (یعنی معادله ۲-۲۴) را با هسته‌های تاری مختلف، فاکتورهای مقیاس و سطوح نویز از طریق یک مدل واحد مدیریت کند. از سوی دیگر، مشابه روش‌های مبتنی بر یادگیری، USRNet را می‌توان به صورت سرتاسری آموزش داد تا اثربخشی و کارایی را تضمین کند.

برای دستیابی به این هدف، ابتدا تابع انرژی مبتنی بر مدل را از طریق یک الگوریتم تقسیم نیمه درجه دوم بازگشایی می‌کنیم. به همین ترتیب، می‌توانیم استنتاجی به دست آوریم که به طور متناوب

بین حل دو مشکل فرعی، یکی مربوط به یک عبارت داده و دیگری به یک عبارت پیش‌فرض قبلی، تغییر می‌کند. سپس با جایگزین کردن راه‌حل‌های دو مشکل فرعی با مازول‌های عصبی، استنتاج را به عنوان یک شبکه عمیق در نظر می‌گیریم. از آنجایی که دو مشکل فرعی به ترتیب با اعمال دانش سازگاری تخریب و تضمین اطلاعات قبلی حذف‌کننده مطابقت دارند، USRNet با تخریب صریح و محدودیت‌های قبلی کاملاً اصولی دارد، که یک مزیت متمایز نسبت به روش‌های SISR مبتنی بر یادگیری موجود است. شایان ذکر است که از آنجایی که USRNet شامل یک فرا پارامتر برای هر مشکل فرعی است، شبکه شامل یک مازول اضافی برای تولید فراپارامتر است. علاوه بر این، به منظور کاهش تعداد پارامترها، تمام مازول‌های قبلی معماری مشابه و پارامترهای مشابهی دارند.

۲-۴-۱ الگوریتم منعطف SISR

اگرچه روش‌های SISR مبتنی بر CNN به موفقیت چشمگیری در مدیریت تخریب دو مکعبی^{۶۷} دست یافته‌اند، استفاده از آنها برای مقابله با سایر مدل‌های تخریب کاربردی‌تر ساده نیست. به منظور عملی بودن، ترجیح داده می‌شود که یک ابر حل‌کننده انعطاف‌پذیر طراحی شود که سه عامل کلیدی، یعنی ضریب مقیاس^{۶۸}، هسته تاری^{۶۹} و سطح نویز^{۷۰} را در نظر بگیرد. چندین روش برای مقابله با تخریب دو مکعبی با فاکتورهای مقیاس مختلف از طریق یک مدل ارائه شده است. برای مقابله انعطاف‌پذیر با یک تصویر LR، برخی روش‌های پیشنهادی هسته تاری کاهش بعد یافته PCA را به عنوان ورودی در نظر می‌گیرند. با این حال، این روش‌ها به هسته‌های تاری گاوسی محدود می‌شوند.

شاید انعطاف‌پذیرترین کارهای مبتنی بر CNN که می‌توانند هسته‌های مختلف تاری، فاکتورهای مقیاس و سطوح نویز را مدیریت کنند، روش‌های عمیق plug-and-play هستند. ایده اصلی چنین روش‌هایی این است که CNN آموخته شده را قبل از راه حل تکراری تحت چارچوب MAP وصل کنیم. متأسفانه، اینها اساساً روش‌های مبتنی بر مدل هستند که از بار محاسباتی بالایی رنج می‌برند و شامل ابرپارامترهای انتخابی دستی هستند. چگونگی طراحی یک مدل آموزش‌پذیر سرتاسری به‌گونه‌ای که بتوان با تکرارهای کمتر به نتایج بهتری دست یافت، هنوز بررسی نشده است. در حالی که بازسازی تصویر کور مبتنی بر یادگیری اخیراً توجه قابل توجهی را به خود جلب کرده است، ما متذکر می‌شویم که این کار بر SISR غیر کور متمرکز است که فرض می‌کند تصویر LR، هسته تاری و سطح نویز از قبل

^{۶۷} bicubic degradation

^{۶۸} scale factor

^{۶۹} blur kernel

^{۷۰} noise level

شناخته شده است. در واقع، SISR غیر کور هنوز یک جهت تحقیقاتی فعال است. ابتدا، هسته تاری و سطح نویز را می‌توان تخمین زد، یا بر اساس اطلاعات دیگر (به عنوان مثال، تنظیمات دوربین) شناخته می‌شود. دوم، کاربران می‌توانند با تنظیم کرنل تاری و سطح نویز، اولویت وضوح و صافی را کنترل کنند. سوم، SISR غیر کور می‌تواند یک گام میانی برای حل SISR کور باشد.

۲-۴-۲ معماری شبکه USRNet

با توجه به چارچوب MAP، تصویر HR را می‌توان با به حداقل رساندن تابع انرژی زیر تخمین زد:

$$E(x) = \frac{1}{2\sigma^2} \|y - (x \otimes k) \downarrow_s\|^2 + \lambda \Phi(x) \quad (25-2)$$

که در آن $\frac{1}{2\sigma^2} \|y - (x \otimes k) \downarrow_s\|^2$ عبارت وفاداری داده، $\Phi(x)$ عبارت پیش‌فرض قبلی و λ یک پارامتر متعادل‌سازی است. به منظور به دست آوردن یک استنتاج آشکار برای معادله ۲-۲۵، الگوریتم تقسیم نیمه درجه دوم^{۷۱} (HQS) به دلیل سادگی و همگرایی سریع در بسیاری از کاربردها انتخاب شده است. HQS معادله ۲-۲۵ را با معرفی یک متغیر کمکی z که منجر به معادل تقریبی زیر می‌شود:

$$E_\mu(x, z) = \frac{1}{2\sigma^2} \|y - (z \otimes k) \downarrow_s\|^2 + \lambda \Phi(x) + \frac{\mu}{2} \|z - x\|^2 \quad (26-2)$$

که در آن μ پارامتر جریمه است. چنین مسئله‌ای را می‌توان با حل مکرر زیرمسائل برای x و z به جواب رساند:

$$\begin{cases} z_k &= \arg \min_z \|y - (z \otimes k) \downarrow_s\|^2 + \mu \sigma^2 \|z - x_{k-1}\|^2 \\ x_k &= \arg \min_x \frac{\mu}{2} \|z_k - x\|^2 + \lambda \Phi(x) \end{cases} \quad (27-2)$$

با توجه به معادله ۲-۲۷، μ باید به اندازه کافی بزرگ باشد به طوری که x و z تقریباً برابر با نقطه ثابت باشند. گرچه این می‌تواند به همگرایی آهسته منجر شود. برای راحتی، μ در تکرار k -ام با μ_k نشان داده شده است. برای حل معادله ۲-۲۷، تبدیل فوری سریع^{۷۲} (FFT) را می‌توان با فرض اینکه کانولوشن با

^{۷۱} half-quadratic splitting

^{۷۲} fast Fourier transform

شرایط مرزی دایره‌ای انجام می‌شود، استفاده کرد:

$$z_k = \mathcal{F}^{-1} \left(\frac{1}{\alpha_k} \left(d - \overline{\mathcal{F}(k)} \odot_s \frac{(\mathcal{F}(k)d) \downarrow_s}{(\overline{\mathcal{F}(k)}\mathcal{F}(k)) \downarrow_s + \alpha_k} \right) \right) \quad (28-2)$$

به طوری که d به صورت زیر تعریف شده:

$$d = \overline{\mathcal{F}(k)}\mathcal{F}(y \uparrow_s) + \alpha_k \mathcal{F}(x_{k-1})$$

با $a_k \triangleq \mu_k \sigma^2$ و در جایی که $\mathcal{F}(\cdot)$ و $\mathcal{F}^{-1}(\cdot)$ نشان دهنده FFT و معکوس FFT است، $\overline{\mathcal{F}(\cdot)}$ نشان دهنده مزدوج مختلط $\mathcal{F}(\cdot)$ است، \odot_s نشان دهنده عملگر پردازش بلوک متمایز با ضرب عنصری، به عنوان مثال، اعمال ضرب عنصری بر روی بلوک‌های متمایز $s \times s$ از $\mathcal{F}(k)$ است. \downarrow_s نشان دهنده نمونه‌برداری با وضوح پایین بلوک متمایز است، یعنی میانگین‌گیری بلوک‌های متمایز $s \times s$ ، \uparrow_s نشان دهنده s-fold upsampler استاندارد است، یعنی نمونه برداری از اندازه فضایی با پر کردن ورودی‌های جدید با صفر. به ویژه قابل توجه است که معادله 28-2 همچنین برای حالت خاص رفع تاری وقتی $s = 1$ است کار می‌کند. برای حل معادله 27-2، مشخص است که از دیدگاه بیزی، در واقع با یک مشکل حذف نویز با سطح نویز $\beta_k \triangleq \sqrt{\frac{\lambda}{\mu_k}}$ مطابقت دارد.

هنگامی که بهینه‌سازی بازگشایی شد، گام بعدی طراحی شبکه وضوح فوق العاده بازگشایی (USRNet) است. از آنجایی که بهینه‌سازی بازگشایی شده عمدتاً شامل حل تکراری یک زیرمسئله داده (یعنی بخش اول معادله 27-2) و یک زیرمسئله قبلی (یعنی بخش دوم معادله 27-2) است، USRNet باید بین یک ماژول داده D و یک ماژول پیش‌فرض قبلی P جایگزین شود. علاوه بر این، از آنجایی که راه‌حل‌های زیرمسئله‌ها نیز فرآیندهای α_k و β_k را به ترتیب ورودی می‌گیرند، یک ماژول فرآیندهای \mathcal{H} اضافه‌تر به USRNet معرفی می‌شود. شکل 9-2 معماری کلی USRNet را با تکرارهای K نشان می‌دهد، که در آن K به طور تجربی روی 8 تنظیم شده است تا بین سرعت-دقت تعادل برقرار شود. در ادامه، جزئیات بیشتری در مورد این سه ماژول ارائه شده است.

ماژول داده D : ماژول داده نقش معادله 28-2 را ایفا می‌کند که راه حل بسته زیرمسئله داده است. به طور شهودی، هدف آن یافتن یک تصویر HR واضح‌تر است که ترکیبی وزنی از عبارت داده‌ای $\|y - (z \otimes k) \downarrow_s\|^2$ و عبارت منظم‌سازی درجه دوم $\|z - x_{k-1}\|^2$ با فرآیندهای سبک-سنگینی α_k را به حداقل می‌رساند. از آنجایی که عبارت داده با مدل تخریب مطابقت دارد، بنابراین ماژول داده نه تنها

از مزیت استفاده از ضریب مقیاس s و محو کردن هسته k به عنوان ورودی برخوردار است، بلکه یک محدودیت تخریب را نیز بر راه حل تحمیل می‌کند. در واقع، طراحی دستی چنین ماژول چند ورودی ساده اما مفید دشوار است. برای اختصار، معادله ۲-۲۸ به صورت زیر بازنویسی می‌شود:

$$z_k = D(x_{k-1}, s, k, y, \alpha_k) \quad (29-2)$$

توجه داشته باشید که x با درون‌یابی y با ضریب مقیاس s از طریق ساده‌ترین درون‌یابی همسایه، مقداردهی اولیه می‌شود. لازم به ذکر است که معادله ۲-۲۹ شامل هیچ پارامتر قابل آموزش نیست، که به نوبه خود به دلیل جداسازی کامل بین عبارت داده و عبارت قبلی، به تعمیم پذیری بهتری منجر می‌شود. برای پیاده‌سازی، ما از PyTorch استفاده می‌کنیم که در آن عملگرهای اصلی FFT و FFT معکوس می‌توانند به ترتیب توسط `torch.rfft` و `torch.irfft` پیاده‌سازی شوند.

ماژول پیش‌فرض قبلی \mathcal{P} : هدف ماژول قبلی بدست آوردن تصویر HR تمیزتر x_k با عبور z_k از یک حذف نویز با سطح نویز β_k است. ما یک حذف کننده نویز عمیق CNN را پیشنهاد می‌کنیم که سطح نویز را به عنوان ورودی می‌گیرد.

$$x_k = \mathcal{P}(z_k, \beta_k) \quad (30-2)$$

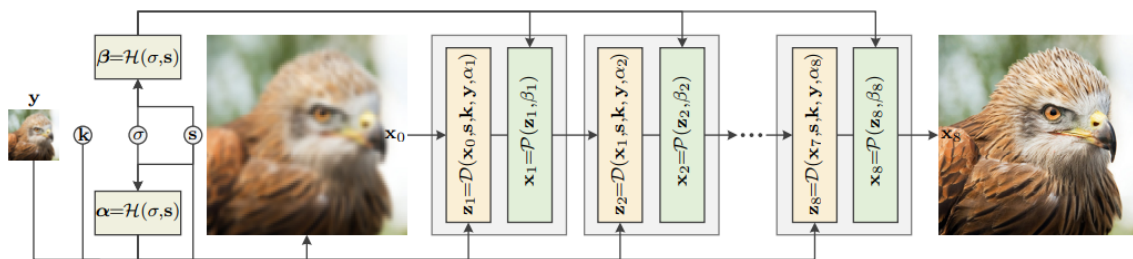
حذف کننده نویز پیشنهادی، یعنی ResUNet، بلوک‌های باقی‌مانده‌ها در U-Net یکپارچه می‌کند. U-Net به طور گسترده‌ای برای نقشه‌برداری تصویر به تصویر استفاده می‌شود، در حالی که ResNet محبوبیت خود را مدیون آموزش سریع و ظرفیت زیاد آن با بسیاری از بلوک‌های باقی‌مانده است. ResUNet نقشه به هم پیوسته z_k و سطح نویز را به عنوان ورودی می‌گیرد و تصویر حذف شده x_k را خروجی می‌دهد. با انجام این کار، ResUNet می‌تواند سطوح مختلف نویز را از طریق یک مدل کنترل کند، که به طور قابل توجهی تعداد کل پارامترها را کاهش می‌دهد. پس از تنظیم مشترک U-Net، ResUNet شامل چهار مقیاس است که هر کدام یک ارتباط پرش هویت بین عملیات کاهش مقیاس و افزایش مقیاس دارند. به طور مشخص تعداد کانال‌ها در هر لایه از مقیاس اول تا مقیاس چهارم به ترتیب ۶۴، ۱۲۸، ۲۵۶ و ۵۱۲ است. برای عملیات کاهش مقیاس و افزایش مقیاس، پیچیدگی گام‌های 2×2 (SConv) و پیچیدگی 2×2 انتقالی (TConv) به ترتیب اتخاذ می‌شوند. توجه داشته باشید که هیچ تابع فعال‌سازی با لایه‌های SConv و TConv و همچنین اولین و آخرین لایه‌های کانولوشن دنبال نمی‌شود. به منظور به ارث بردن

امتیازات ResNet، گروهی از ۲ بلوک باقیمانده در مقیاس کوچک و افزایش مقیاس هر مقیاس اتخاذ می‌شوند. همانطور که پیشنهاد شد، هر بلوک باقیمانده از دو لایه پیچشی 3×3 با فعال‌سازی ReLU در وسط و یک اتصال پرش هویت که به خروجی آن خلاصه می‌شود، تشکیل شده است.

ماژول هایپرپارامتر \mathcal{H} : ماژول هایپرپارامتر به عنوان یک نوار اسلاید برای کنترل خروجی‌های ماژول داده و ماژول قبلی عمل می‌کند. به عنوان مثال، با افزایش α_k ، راه حل z_k به تدریج به x_{k-1} نزدیک می‌شود. طبق تعریف α_k ، β_k و σ توسط μ_k تعیین می‌شود، در حالی که β_k به λ و μ_k بستگی دارد. اگرچه یادگیری λ و μ_k ثابت امکان پذیر است، اما ما استدلال می‌کنیم که اگر λ و μ_k با دو عنصر کلیدی، یعنی ضریب مقیاس s و سطح نویز σ ، که بر درجه تأثیر می‌گذارند، تغییر کنند، می‌توان افزایش عملکرد را به دست آورد. بد اخلاقی فرض کنید $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_K]$ و $\beta = [\beta_1, \beta_2, \dots, \beta_K]$ ، ما از یک ماژول واحد برای پیش‌بینی α و β استفاده می‌کنیم

$$[\alpha, \beta] = \mathcal{H}(\sigma, s) \quad (31-2)$$

ماژول هایپرپارامتر از سه لایه کاملاً متصل با ReLU به عنوان دو تابع فعال‌سازی اول و Softplus به عنوان آخرین لایه تشکیل شده است. تعداد گره‌های پنهان در هر لایه ۶۴ عدد است. با توجه به اینکه α_k و β_k باید مثبت باشند و معادله ۲-۳۱ باید از تقسیم با α_k بسیار کوچک اجتناب شود، لایه Softplus خروجی با افزودن اضافی $e - 6$ دنبال می‌شود. ما نشان خواهیم داد که چگونه ضریب مقیاس و سطح نویز بر پارامترهای فوق در ثانیه تأثیر می‌گذارد.



شکل ۲-۹: معماری کلی USRNet پیشنهادی با $K = 8$ تکرار. USRNet می‌تواند به‌طور انعطاف‌پذیری تخریب کلاسیک (یعنی معادله ۲-۲۴) را از طریق یک مدل مدیریت کند زیرا تصویر LR y ، ضریب مقیاس s ، کرنل تاری k و سطح نویز σ را به عنوان ورودی می‌گیرد. به‌طور خاص، USRNet از سه ماژول اصلی تشکیل شده است، از جمله ماژول داده D که تخمین HR را واضح‌تر می‌کند، ماژول پیش‌فرض قبلی P که تخمین HR را تمیزتر می‌کند و ماژول فرایارامتر \mathcal{H} که خروجی‌های D و P را کنترل می‌کند.

۵-۲ رفع نویز تصاویر (Image denoising)

روش‌های بازگشایی عمیق موجود، یک الگوریتم بهینه‌سازی را با تعداد ثابتی از مراحل باز می‌کنند و از شبکه‌های عصبی کانولوشنال (CNN) برای یادگیری پیشین‌های مبتنی بر داده استفاده می‌کنند. با این حال، عملکرد آنها به دو دلیل اصلی محدود است. اولاً، پیش‌فرض‌هایی که در فضای ویژگی عمیق آموخته می‌شوند باید در هر مرحله تکرار به فضای تصویر تبدیل شوند، که عمق CNN ها را محدود می‌کند و از بهره‌برداری CNN از اطلاعات متنی جلوگیری می‌کند. ثانیاً، روش‌های موجود فقط پیشین‌های عمیق را در مقیاس تکی با وضوح کامل یاد می‌گیرند، بنابراین از مزایای زمینه چند مقیاسی در مواجهه با نویز سطح بالا چشم‌پوشی می‌کنند. برای پرداختن به این مسائل، نویسندگان مقاله "Deep unfolding multi-scale regularizer network for image denoising" [۶] به صراحت فرآیند حذف نویز تصویر را در فضای ویژگی عمیق در نظر می‌گیرند و شبکه تنظیم‌کننده چند مقیاسی بازگشایی عمیق (DUMRN) را برای حذف نویز تصویر پیشنهاد می‌کنند. هسته DUMRN ماژول حذف نویز مبتنی بر ویژگی^{۷۳} (FDM) است که به طور مستقیم نویز را در فضای ویژگی عمیق حذف می‌کند. در هر FDM، یک بلوک تنظیم‌کننده چند مقیاسی برای یادگیری اطلاعات عمیق قبلی از ویژگی‌های چند وضوحی ایجاد می‌شود. نویسندگان DUMRN را با چیدن دنباله ای از FDM ها می‌سازند و آن را به صورت سرتاسری آموزش می‌دهند. نتایج تجربی روی معیارهای مصنوعی و دنیای واقعی نشان می‌دهد که DUMRN در مقایسه با روش‌های پیشرفته عملکرد مطلوبی دارد.

به دلیل ماهیت بد وضع مسئله حذف نویز تصویر، بسیاری از روش‌های مرسوم از عبارات پیش‌فرض قبلی تصویر بر اساس آمار تصاویر طبیعی استفاده می‌کنند. علیرغم پیشرفت قابل توجه آنها، این روش‌های مبتنی بر مدل معمولاً تصاویر تمیز پنهان را با حل مسائل بهینه سازی پیچیده بازسازی می‌کنند که کاربرد عملی آنها را محدود می‌کند.

سایر روش‌ها برای دستیابی به عملکرد بالا، انعطاف پذیری و کارایی را قربانی می‌کنند. با توسعه شبکه‌های عصبی پیچشی عمیق (CNN)، روش‌های مبتنی بر یادگیری بسیاری برای حذف نویز تصویر پیشنهاد شده‌اند. با توانایی نمایش قدرتمند CNN های عمیق، این روش‌ها با فاصله زیادی از روش‌های سنتی مبتنی بر مدل بهتر عمل می‌کنند. با این حال، اکثر روش‌های مبتنی بر یادگیری مستقیماً نداشت بین جفت‌های تصویر نویزدار و تمیز را بدون در نظر گرفتن مدل فیزیکی فرآیند نویز یاد می‌گیرند، که باعث می‌شود آنها کارآمدتر اما کمتر قابل تفسیر نسبت به روش‌های مبتنی بر مدل باشند.

^{۷۳}feature-based denoising module

با بهره‌گیری از روش‌های مبتنی بر مدل و مبتنی بر یادگیری، بسیاری از روش‌های بازگشایی عمیق، روش‌های بهینه‌سازی استاندارد را در CNN های عمیق ترکیب می‌کنند. آنها مسئله حذف نویز تصویر را از طریق الگوریتم‌های بهینه‌سازی مختلف آشکار می‌کنند (به عنوان مثال، گرادیان کاهشی)، و عبارت منظم‌سازی را با استفاده از CNN های عمیق پیاده‌سازی می‌کنند، که می‌توانند به طور ضمنی پیش‌فرض‌های عمیق را در فضای ویژگی یاد بگیرند. با ادغام محدودیت تخریب تصویر در CNN، روش‌های بازگشایی عمیق کارایی را حفظ کرده و تفسیرپذیری یادگیری عمیق را بهبود می‌بخشند. با این حال، عملکرد روش‌های آشکارسازی عمیق فعلی به دو دلیل اصلی هنوز محدود است. اولاً، در هر مرحله تکرار از روش‌های بازگشایی عمیق، نویز در فضای تصویر استاندارد حذف می‌شود، اما اولویت‌های عمیق در فضای ویژگی‌ها آموخته می‌شوند. تبدیل از فضای ویژگی عمیق به فضای تصویر، عمق و میدان دریافت CNN ها را محدود می‌کند، که از استخراج وابستگی‌های غیر محلی درون تصاویر توسط CNN ها جلوگیری می‌کند. ثانیاً، روش‌های بازگشایی عمیق موجود، فقط پیش‌فرض‌های عمیق را در وضوح کامل اصلی یاد می‌گیرند، و بنابراین نمی‌توانند به طور موثر اطلاعات زمینه‌ای مکانی را ضبط کنند و لبه‌های واضح را برای تصاویری که از نویز سنگین رنج می‌برند بازیابی کنند.

در این مقاله، نویسندگان شبکه تنظیم‌کننده چند مقیاسی عمیق (DUMRN) را برای ادغام دقیق‌تر روش‌های مبتنی بر مدل و مبتنی بر یادگیری پیشنهاد می‌کنند. برای کاهش تعداد تبدیل‌های فضا و بهبود جریان اطلاعات در داخل شبکه، به صراحت فرآیند حذف نویز تصویر را در فضای ویژگی عمیق در نظر می‌گیرند و یک ماژول حذف نویز مبتنی بر ویژگی (FDM) را بر اساس مدل تخریب تصویر پیشنهاد می‌کنند. با تقلید از فرآیند بهینه‌سازی گرادیان کاهشی، دنباله‌ای از FDM ها برای ساخت DUMRN روی هم چیده می‌شوند، به طوری که می‌توانند یک CNN عمیق با یک میدان پذیرنده بزرگ به دست آورند و آن را به روشی سرتاسری آموزش دهند. در هر FDM، یک بلوک تنظیم‌کننده چند مقیاسی^{۷۴} (MSRB) برای یادگیری اطلاعات قبلی عمیق از ویژگی‌های چند رزولوشن می‌سازند، که می‌تواند جزئیات محلی را با وضوح بالا و اطلاعات متنی در مقیاس بزرگ را با وضوح پایین ثبت کند.

^{۷۴} multi-scale regularizer block

۲-۵-۱ مازول حذف نویز مبتنی بر ویژگی (FDM)

از دیدگاه بیزی، حداکثرسازی یک مدل پسین (MAP) برای نویز زدایی می‌تواند به صورت زیر فرموله شود:

$$\hat{x} = \arg \min_x \frac{1}{2} \|y - x\|_2^2 + \lambda \phi(x) \quad (۳۲-۲)$$

که در آن اولین عبارت وفاداری تضمین می‌کند که پاسخ \hat{x} مطابق با فرآیند تخریب، $\phi(x)$ عبارت منظم‌سازی مرتبط با پیش‌فرض‌های قبلی، و λ پارامتر منظم‌سازی است که متعادل‌سازی بین این عبارت‌ها را کنترل می‌کند. معادله ۲-۳۲ را می‌توان با الگوریتم‌های بهینه‌سازی مختلف، مانند گرادیان کاهشی حل کرد. در اینجا ما از روش نزول گرادیان تکانه‌ای^{۷۵} (MGD) به دلیل سادگی و اثربخشی آن استفاده می‌کنیم. بنابراین، \hat{x} را می‌توان از طریق تکرار در معادلات زیر به دست آورد:

$$x^{t+1} = x^t - \alpha^t (x^t - y) - \lambda \nabla \phi^t(x^t) + \beta^t (x^t - x^{t-1}) \quad (۳۳-۲)$$

$$y = x^\circ \quad (۳۴-۲)$$

که در آن t نشان‌دهنده شاخص مرحله، α^t اندازه گام، $\nabla \phi^t(\cdot)$ نشان‌دهنده گرادیان $\phi(\cdot)$ ، β_t پارامتر تکانه در مرحله t است، و $(x^t - x^{t-1})$ جمله تکانه است. به منظور یادگیری پیش‌فرض‌های عمیق تصویر، روش‌های بازگشایی عمیق از CNN برای محاسبه گرادیان منظم‌ساز $\nabla \phi_t(\cdot)$ استفاده می‌کنند، که به طور ضمنی یک پیش‌فرض عمیق از تصویر را ارائه می‌دهد. سپس رویه تکراری در معادله ۲-۳۳ با تعداد ثابتی از مراحل T بازگشایی می‌شود. با این حال، در هر مرحله تکرار، نویز در x^t در فضای تصویر استاندارد حذف می‌شود در حالی که $\nabla \phi^t(\cdot)$ در فضای ویژگی عمیق آموخته می‌شود. نقشه برداری از فضای ویژگی عمیق به فضای تصویر، عمق و اندازه میدان دریافتی CNN ها را محدود می‌کند، و روش‌های بازگشایی را برای گرفتن وابستگی‌های غیر محلی در داخل تصاویر دشوار می‌کند.

به منظور استفاده کامل از CNN های عمیق و بهبود جریان اطلاعات در شبکه، ما به صراحت حذف نویز در فضای ویژگی عمیق را در نظر می‌گیریم. به طور خاص، ابتدا از یک استخراج‌کننده ویژگی $f(\cdot)$ برای نگاشت تصویر x به فضای ویژگی استفاده می‌کنیم. سپس می‌توانیم از معادله ۲-۳۳ برای حذف

^{۷۵} momentum gradient descent

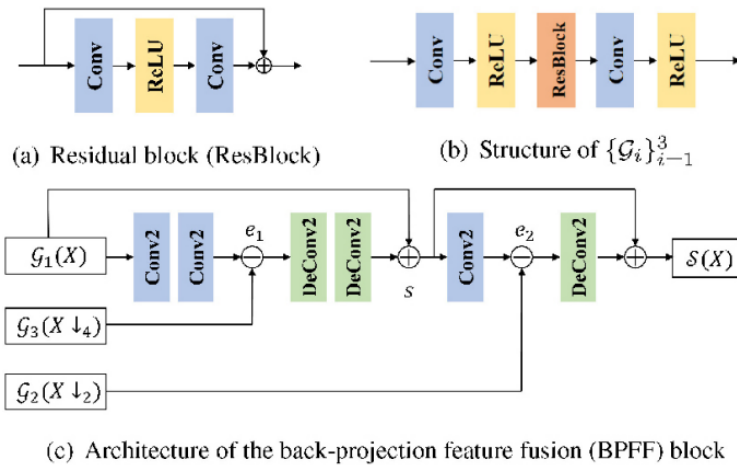
مکرر $f(x)$ بدون نگاشت آن در فضای تصویر استفاده کنیم. به منظور مقیاس بندی شرایط گرادیان و تکانه به صورت تطبیقی، اندازه گام α^t و پارامتر تکانه β^t را به ترتیب با $A^t(\cdot)$ و $B^t(\cdot)$ جایگزین می‌کنیم. اجازه می‌دهیم $S(\cdot)$ جایگزین $\nabla\phi^t(\cdot)$ شود تا شیب منظم‌ساز را جایگزین کند. $S(\cdot)$ وزن λ را جذب می‌کند. $S(\cdot)$ به طور ضمنی نقش پیش‌فرض قبلی عمیق را انجام می‌دهد. به طور کلی، مازول حذف نویز مبتنی بر ویژگی t (FDM) به صورت زیر فرموله شده است:

$$X^{t+1} = X^t - A^t (X^t - Y + S_t(X^t)) + B^t (X^t - X^{t-1}) \quad (35-2)$$

$$Y = f(y) = f(x^\circ) = X^\circ \quad (36-2)$$

پس از تکرار T ، از یک بازساز تصویر $g(\cdot)$ برای بازسازی تصویر حذف شده نهایی از X^T استفاده می‌کنیم: $\hat{x} = g(X^T)$

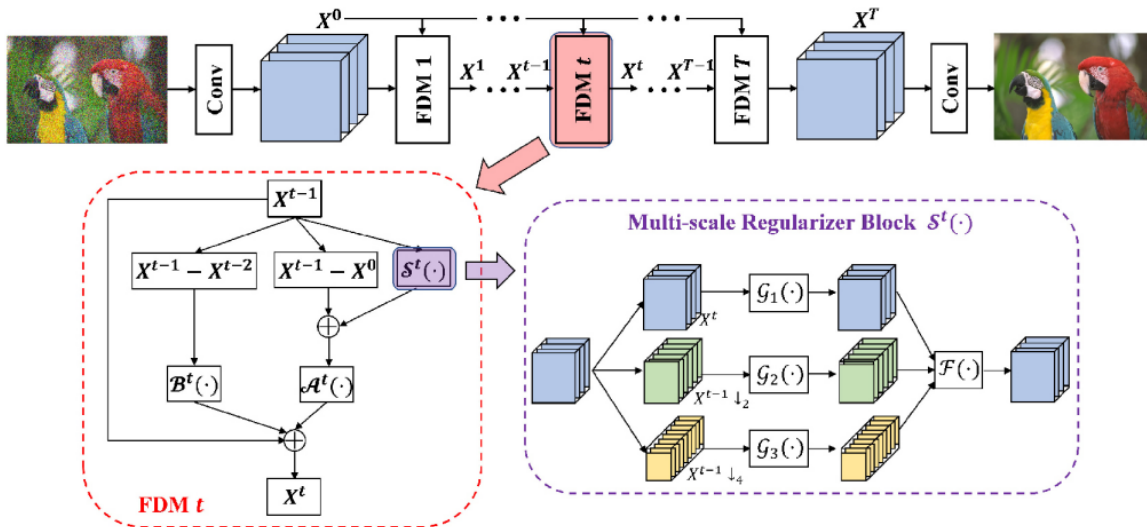
با در نظر گرفتن اثربخشی بلوک باقیمانده (ResBlock) در وظایف بینایی سطح پایین، ما از یک ResBlock برای پیاده سازی $A^t(\cdot)$ و $B^t(\cdot)$ استفاده می‌کنیم. ساختار ResBlock در شکل ۱۰-۲ (a) نشان داده شده است: هر ResBlock شامل دو لایه 3×3 کانولوشن و یک تابع فعال‌سازی ReLU است. برای انعطاف‌پذیری، استخراج‌کننده ویژگی f و بازسازی‌کننده تصویر $g(\cdot)$ را با استفاده از دو لایه کانولوشن با هسته‌های 3×3 قابل یادگیری پیاده‌سازی می‌کنیم.



شکل ۱۰-۲: (a) ساختار بلوک باقیمانده (ResBlock). (b) $\{G_i(\cdot)\}_{i=1}^3$ در FDM یک معماری مشترک از یک CNN با پارامترهای مختلف برای یادگیری به اشتراک می‌گذارد. (c) ساختار بلوک BPFF و DeConv2 و Conv2. به ترتیب یک لایه کانولوشن و لایه بازگردانی کانولوشن با گام ۲ را نشان می‌دهند.

۲-۵-۲ معماری شبکه DUMRN

معماری کلی شبکه تنظیم کننده چند مقیاسی بازگشایی عمیق پیشنهادی ما (DUMRN) در شکل ۱۱-۲ نشان داده شده است. اگر y و x به ترتیب تصویر نویزدار ورودی و تصویر واقعی زمین مربوطه را نشان دهند. استخراج کننده ویژگی f ابتدا ویژگی‌ها را از y استخراج می‌کند. ویژگی‌های استخراج شده Y به عنوان مقدار اولیه X^0 برای فرآیند حذف نویز از ویژگی تنظیم می‌شود. FDM T برای حذف نویز در فضای ویژگی روی هم چیده شده‌اند: ما X^t را از طریق T مرحله با استفاده از معادله ۲-۳۵ بروز می‌کنیم. در نهایت، ما از بازسازی تصویر g برای بازسازی تصویر نهایی از ویژگی حذف شده X^T استفاده می‌کنیم. DUMRN با به حداقل رساندن تفاوت بین تصاویر نویزدار ورودی و همتایان واقعی مرجع مربوطه بهینه شده است. برای ارزیابی اثربخشی شبکه پیشنهادی، ما همان تابع از خطا L_2 را مانند کارهای قبلی انجام شده در این زمینه، اتخاذ می‌کنیم. با توجه به مجموعه داده آموزشی $\{y_i, x_i\}_{i=1}^N$ ، که در آن N تعداد جفت‌های پیچ آموزشی است، پارامترهای بهینه را با کمینه کردن تابع هدف در معادله ۲-۳۶ $L(\Theta) = \frac{1}{N} \sum_{i=1}^N \|f_{DUMRN}(y_i) - x_i\|^2$ بدست می‌آوریم. که در آن $f_{DUMRN}(\cdot)$ نشان دهنده تابع DUMRN است، و Θ نشان دهنده تمام پارامترهای قابل یادگیری در DUMRN است.



شکل ۱۱-۲: معماری DUMRN. اولین استخراج لایه کانولوشن X^0 از ورودی نویزدار را مشخص می‌کند، و سپس ماژول‌های حذف نویز مبتنی بر ویژگی T (FDM) برای حذف نویز در فضای ویژگی عمیق روی هم چیده می‌شوند. ساختار FDM بر اساس الگوریتم گرادین کاهشی تکانه است. با بهره‌مندی از FDM، نیازی به تبدیل ویژگی‌های عمیق به فضای تصویر در هر مرحله وجود ندارد. آخرین لایه کانولوشن، ویژگی حذف شده X^T را به یک تصویر تمیز تبدیل می‌کند. یک بلوک تنظیم کننده چند مقیاسی $S(\cdot)$ اطلاعات عمیق قبلی را از ویژگی‌ها در وضوح‌های مختلف می‌آموزد. \downarrow_k نشان دهنده نمونه برداری پایین با ضریب مقیاس بندی k است.

فصل سوم

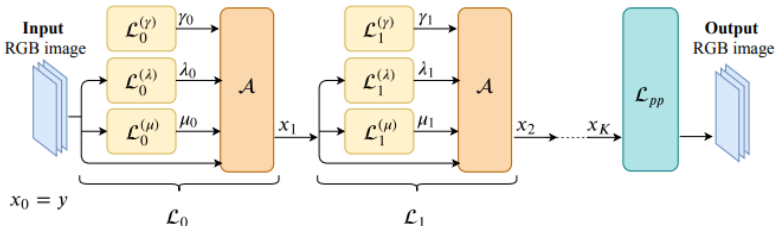
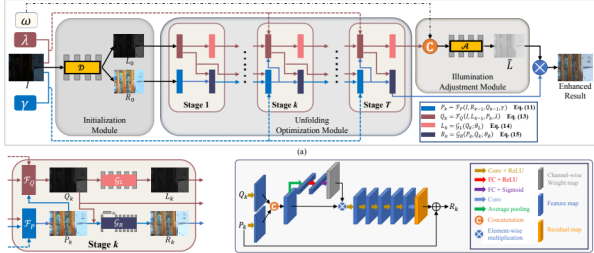
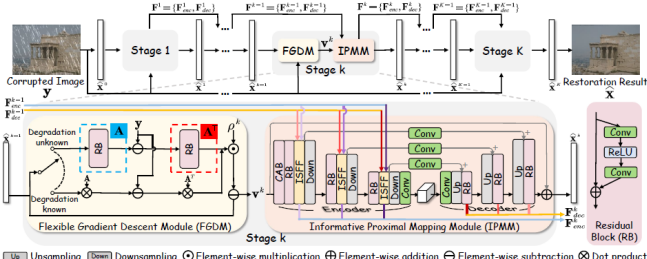
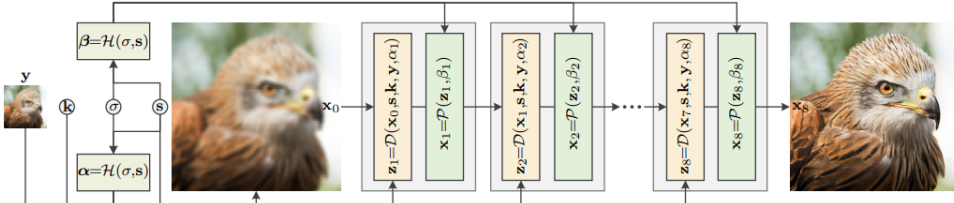
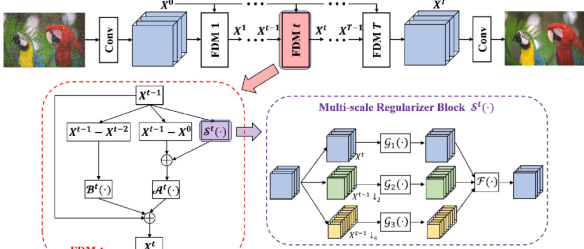
نتیجه گیری

در این پژوهش، ما به بررسی و تحلیل عمیق شبکه‌های بازگشایی عمیق پرداختیم که به منظور بهبود کیفیت تصاویر در موقعیت‌های مختلف مانند رفع تاری، بهبود تصاویر کم‌نور، ترمیم تصاویر، افزایش وضوح تصاویر و رفع نویز تصاویر به کار می‌روند. معماری‌های متنوعی از این شبکه‌ها مورد بررسی قرار گرفت که هر یک دارای ویژگی‌ها و کاربردهای منحصر به فردی بودند. هریک الگوریتم و مدلی را انتخاب و آن را بازگشایی کردند و بر طبق این بازگشایی معماری شبکه عصبی بازگشایی عمیق خود را گسترش دادند.

۳-۱ جدول مقایسه نهایی

در ادامه برای مقایسه این روش‌ها در یک نگاه به صورت کلی، اطلاعات هریک در جدول ۳-۱ جمع‌آوری شده است. در کارهای آینده نیز می‌توانید با انتخاب الگوریتم‌ها و مدل‌های دیگر از بین روش‌های مبتنی بر مدل و مبتنی بر متغیر، و بازگشایی و ترکیب آنها با شبکه‌های عصبی عمیق به گسترش این حوزه بپردازید.

جدول ۱-۳

معماری کلی شبکه	نام شبکه
	iRestNet
الگوریتم بازگشایی شده نقطه داخلی نزدیک	فرموله‌بندی مسئله بد وضع معکوس $\min_{x \in \mathbb{R}^n} f(Hx, y) + \lambda \mathcal{R}(x) + \mu \mathcal{B}(x)$
	URetinex-Net
الگوریتم بازگشایی شده مبتنی بر رتینکس	فرموله‌بندی مسئله بد وضع معکوس $\min_{P, Q, R, L} \ I - P \cdot O\ _F + \alpha \Phi(R) + \beta \Psi(L)$
	DGUNet
الگوریتم بازگشایی شده گردایان کاهشی نزدیک	فرموله‌بندی مسئله بد وضع معکوس $x = \operatorname{argmin}_x \frac{1}{p} \ y - Ax\ _p + \lambda J(x)$
	USRNet
الگوریتم بازگشایی شده ابر وضوح تک تصویر معطف	فرموله‌بندی مسئله بد وضع معکوس $E(x) = \frac{1}{\sqrt{\sigma^2}} \ y - (x \otimes k)\ _{\downarrow s} + \lambda \Phi(x)$
	DUMRN
الگوریتم بازگشایی شده حذف نویز مبتنی بر ویژگی	فرموله‌بندی مسئله بد وضع معکوس $E(x) = \frac{1}{\sqrt{\sigma^2}} \ y - (x \otimes k)\ _{\downarrow s} + \lambda \Phi(x)$

منابع و مراجع

- [1] Bertocchi, Carla, Chouzenoux, Emilie, Corbineau, Marie-Caroline, Pesquet, Jean-Christophe, and Prato, Marco. Deep unfolding of a proximal interior point method for image restoration. *Inverse Problems*, 36(3):034005, 2020.
- [2] Wu, Wenhui, Weng, Jian, Zhang, Pingping, Wang, Xu, Yang, Wenhan, and Jiang, Jianmin. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5901–5910, 2022.
- [3] Mou, Chong, Wang, Qian, and Zhang, Jian. Deep generalized unfolding networks for image restoration. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17399–17410, 2022.
- [4] Zhang, Kai, Gool, Luc Van, and Timofte, Radu. Deep unfolding network for image super-resolution. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3217–3226, 2020.
- [5] Ning, Qian, Dong, Weisheng, Shi, Guangming, Li, Leida, and Li, Xin. Accurate and lightweight image super-resolution with model-guided deep unfolding network. *IEEE Journal of Selected Topics in Signal Processing*, 15(2):240–252, 2020.
- [6] Xu, Jingzhao, Yuan, Mengke, Yan, Dong-Ming, and Wu, Tieru. Deep unfolding multi-scale regularizer network for image denoising. *Computational Visual Media*, 9(2):335–350, 2023.

Abstract

Image quality enhancement remains a challenging task in computer vision. Traditional methods, such as model-based and variational-based approaches, have limitations despite their advantages. In recent years, deep unfolding networks (DUNs) have emerged as a novel approach for solving ill-posed inverse problems in image quality enhancement. DUNs combine the strengths of model-based and deep learning methods (deep neural networks), preserving efficiency while improving the interpretability of deep learning.

This paper comprehensively reviews diverse deep unfolding network architectures for image quality enhancement. Firstly, we introduce the concept of deep unfolding and its advantages over traditional methods. Then, we delve into various deep unfolding network architectures, including iRestNet, URetinex-Net, DGUNet, USRNet, MoG-DUN, and DUMRN. Next, we briefly discuss the applications of these networks in various image quality enhancement tasks, such as deblurring, low-light enhancement, restoration, super-resolution, and denoising. Finally, we discuss the challenges and future research opportunities in deep unfolding networks for image quality enhancement.

Key Words:

(Deep Unfolding, Neural Networks, Image Enhancement, Ill-posed Inverse Problems)



**Amirkabir University of Technology
(Tehran Polytechnic)**

Department of Mathematics and Computer Science

B. Sc. Thesis

Hyperparameters Tuning of Deep Unfolding Networks for Image Enhancement

By

Pouria Alimoradpor

Supervisor

Dr. Fatemeh Shakeri

Advisor

Dr. Behzad Najafi

Winter 2024