عنوان مقاله :

UnfoldIR: Rethinking Deep Unfolding Network in Illumination Degradation Image Restoration

کنفرانس : CVPR 2025/2026

بخش اول: مقدمه و مبانی نظریه رتینکس در بازیابی تصاویر

بازیابی تصاویر دچار تخریب نوری (IDIR) مجموعه‌ای از چالش‌های پیچیده در پردازش تصویر را در بر می‌گیرد که هدف آن اصلاح تصاویری است که تحت شرایط نامساعد نوری ثبت شده‌اند .این شرایط نامساعد منجر به بروز پدیده‌های مخربی نظیر کنتراست پایین، از دست رفتن جزئیات ساختاری و غلبه نویز غیریکنواخت در نواحی تاریک تصویر می‌گردد .الگوریتم‌های توسعه‌یافته برای حل این مسئله عموما در دو دسته اصلی طبقه‌بندی می‌شوند: روش‌های مبتنی بر مدل و روش‌های مبتنی بر یادگیری عمیق .روش‌های مبتنی بر مدل به دلیل استفاده از قوانین و قیود ریاضی صریح (مانند توزیع‌های آماری و معادلات دیفرانسیل)، از سطح تفسیرپذیری بالایی برخوردارند، اما در مواجهه با توزیع‌های نوری ناشناخته در تصاویر دنیای واقعی، تعمیم‌پذیری پایینی از خود نشان می‌دهند .در نقطه مقابل، روش‌های یادگیری عمیق از طریق آموزش‌های سرتاسری (End-to-end) بر روی داده‌گان عظیم، تعمیم‌پذیری بسیار بالایی دارند اما به دلیل ماهیت «جعبه سیاه» خود، فاقد تفسیرپذیری و ضمانت‌های ریاضیاتی اثبات‌پذیر هستند .

شبکه‌های بازشده عمیق (Deep Unfolding Networks - DUNs) با رسالت یکپارچه‌سازی مزایای این دو پارادایم معرفی شده‌اند .این شبکه‌ها، الگوریتم‌های بهینه‌سازی تکرارشونده را به لایه‌های یک معماری عصبی نگاشت می‌کنند .با این وجود، عملکردDUN های پیشین در مسئله IDIR به طرز محسوسی پایین‌تر از مدل‌های پیشرو (SOTA) بوده است .معماری UnfoldIR با این استدلال بنا شده است که ضعف یادشده، ریشه در محدودیت‌های ذاتی DUN ندارد؛ بلکه ناشی از عدم فرمول‌بندی یک مدل وظیفه‌محور با قیود اختصاصی، عدم بهره‌گیری از ماژول‌های پیشرفته استخراج ویژگی در فاز بسط شبکه، و فقدان توابع هزینه (Loss Functions) متناسب با ماهیت چندمرحله‌ای این شبکه‌ها است .

فرمول‌بندی پایه تئوری رتینکس (Retinex Theory) هسته مرکزی مدل ریاضی در معماریUnfoldIR ، مبتنی بر تئوری رتینکس است .بر اساس این تئوری فیزیکی–ادراکی، هر تصویر مشاهده‌شده را می‌توان به عنوان حاصل‌ضرب عنصر به عنصر (Hadamard Product) میزان بازتاب ذاتی اشیاء و میزان روشنایی محیط مدل‌سازی کرد .اگر تصویر تخریب‌شده را با I نشان دهیم، معادله تجزیه تصویر به صورت زیر فرمول‌بندی می‌شود :

$$I = R \odot L = \left(R_{HQ} + \hat{R}\right) \odot \left(L_{HQ} + \hat{L}\right)$$

در این معادله:

- مؤلفه R نشان‌دهنده نقشه بازتاب (Reflectance) است که شامل ویژگی‌های ذاتی و مستقل از نور اشیاء، نظیر بافت و رنگ واقعی است .
- مؤلفه L بیانگر نقشه روشنایی (Illumination) است که شدت نور تابیده‌شده به صحنه را مدل می‌کند .
- متغیرهای R_{HQ} و L_{HQ} به ترتیب نمایانگر مؤلفه‌های بازتاب و روشنایی باکیفیت و ایده‌آل نهفته در تصویر هستند .
- مؤلفه $\hat{R}$ نشان‌دهنده اغتشاشات و تخریب‌های بافتی (نظیر نویزهای فرکانس بالا) در مؤلفه بازتاب است .
- مؤلفه $\hat{L}$ بیانگر اعوجاجات نوری و انحرافات رنگی در نقشه روشنایی است .

هدف نهایی الگوریتم پیشنهادی، تقلیل و سرکوب حداکثری اغتشاشات $\hat{R}$ و $\hat{L}$ به منظور استخراج مؤلفه‌های بهینه و بازیابی تصویر نهایی است .

بخش دوم: فرمول‌بندی ریاضی تابع هدف و ترم‌های منظم‌سازی(Regularization)

برای دستیابی به مؤلفه‌های بهینه R و L، فرآیند بازیابی به یک مسئله بهینه‌سازی ریاضیاتی تبدیل می‌شود .تابع هدف (Objective Function) پایه برای این سیستم به فرم زیر تعریف می‌گردد :

$$\mathcal{L}(R, L) = \frac{1}{2}||I - R \odot L||_2^2 + \beta\varphi(R) + \gamma\phi(L)$$

در این معادله، ترم اول $\frac{1}{2}||I - R \odot L||_2^2$ نمایانگر وفاداری به داده (Data Fidelity) است که با استفاده از نرم اقلیدسی (l_2-norm) محاسبه شده و تضمین می‌کند که ترکیب مؤلفه‌های بازسازی‌شده با تصویر ورودی مطابقت داشته باشد .پارامترهای β و γ وزن‌های کنترلی (Trade-off parameters) هستند و ترم‌های $\varphi(R)$ و $\phi(L)$ توابع منظم‌ساز ضمنی هستند که در طول فرآیند آموزش شبکه عمیق به صورت خودکار آموخته می‌شوند .

با این حال، تکیه صرف بر یادگیری ضمنی برای کنترل اغتشاشات تصاویر با تخریب نوری شدید کفایت نمی‌کند. بنابراین، در معماریUnfoldIR ، قیود صریح و دانش‌پایه به تابع هدف اضافه می‌شوند تا رفتار سیستم در برابر نویز و اعوجاج رنگ به صورت قطعی کنترل گردد .تابع هدف نهایی و توسعه‌یافته به صورت زیر تدوین می‌شود :

$$\mathcal{L}(R,L) = \frac{1}{2}||I - R \odot L||_2^2 + \beta\varphi(R) + \gamma\phi(L) + \mu\mathcal{S}\big(\mathcal{A}(I) - \mathcal{A}(R)\big) + \frac{\lambda}{2}||w \odot \nabla L||_2^2$$

تحلیل ترم منظم‌ساز صریح برای مؤلفه بازتاب :(Reflectance) ترم $\mu\mathcal{S}\big(\mathcal{A}(I) - \mathcal{A}(R)\big)$ به طور اختصاصی جهت سرکوب نویزهای تصویربرداری و ارتقای جزئیات بافت به سیستم تزریق شده است .در این ترم، عملگر $\mathcal{S}(\cdot)$ نمایانگر نرم l_1 است که برای وادار کردن سیستم به حفظ ثبات توزیع بافت بین تصویر ورودی (I) و نقشه بازتاب (R) به کار می‌رود .پارامتر μ یک ضریب کنترلی است و عملگر $\mathcal{A}(\cdot)$ با الهام‌گیری از الگوریتم نفوذ ناهمسانگرد پرونا–مالیک (Perona-Malik) مدل‌سازی شده است .فرمول این عملگر برای پیکسل i به شرح زیر است :

$$\mathcal{A}(I_i) = \frac{1}{\bar{\eta}_i}\sum_{j\in\eta_i} c\left(|\nabla I_{i,j}|\right)\nabla I_{i,j}$$

در اینجا، η_i مجموعه پیکسل‌های همسایه در اطراف پیکسل i است و $\bar{\eta}_i$ تعداد این پیکسل‌های همسایه را نشان می‌دهد که روی عدد ۹ تنظیم شده است .نماد ∇ نشان‌دهنده عملگر گرادیان) مانند عملگر سوبل (Sobel - و $|\cdot|$ اندازه (Magnitude) گرادیان را مشخص می‌کند .بخش حیاتی این قید، تابع نفوذ $c(\cdot)$ است که به صورت نمایی تعریف می‌شود :

$$c\big(|\nabla I_{i,j}|\big) = \exp\left[-\left(\frac{|\nabla I_{i,j}|}{s}\right)^2\right]$$

منطق ریاضی و فیزیکی این تابع بدین گونه است که ثابت s حساسیت فرآیند نفوذ را کنترل می‌کند (در مدلUnfoldIR ، s به عنوان یک پارامتر قابل یادگیری در نظر گرفته شده است .(اگر در یک ناحیه از تصویر، مقدار گرادیان $|\nabla I_{i,j}|$ کوچک باشد، سیستم متوجه می‌شود که با یک ناحیه نسبتا هموار (Smooth region)روبرو است؛ در نتیجه، مقدار تابع نمایی به عدد یک نزدیک شده، فرآیند نفوذ تشدید می‌گردد و نویزها به شدت فیلتر می‌شوند .در مقابل، اگر مقدار گرادیان بزرگ باشد، سیستم وجود لبه‌ها و جزئیات مهم بافتی را تشخیص می‌دهد؛ در این حالت مقدار تابع به سمت صفر میل کرده، فرآیند نفوذ تضعیف می‌شود و از محو شدن و تخریب بافت‌های مهم جلوگیری می‌گردد .

تحلیل ترم منظم‌ساز صریح برای مؤلفه روشنایی :(Illumination) ترم $\frac{\lambda}{2}||w \odot \nabla L||_2^2$ برای ایجاد یک قید هموارسازی محلی روی نقشه روشنایی لحاظ شده است تا از پدیده اعوجاج رنگ (Color distortion) جلوگیری کند .در این فرمول، λ متغیر کنترلی است و w به عنوان یک ماتریس وزن‌دهی آگاه از گرادیان به فرم زیر تعریف می‌شود :

$$w = \frac{1}{\exp|\nabla L|}$$

نقش استراتژیک ماتریس وزن‌دهی w این است که با اعمال مجازات (Penalty) معکوس نسبت به اندازه گرادیان، از هموارسازی بیش از حد مرزهای نوری جلوگیری به عمل می‌آورد .به عبارت دیگر، این ماتریس به سیستم اجازه می‌دهد تا روشنایی را در نواحی یکنواخت به خوبی هموار سازد، اما ساختارهای اطلاعاتی مهم و مرزهای نوری را به شدت حفظ کند

بخش سوم: بهینه‌سازی ریاضی (الگوریتم گرادیان پروگزیمال)

برای استخراج بهینه‌ترین مقادیر برای نقشه‌های بازتاب (R) و روشنایی (L) از تابع هدف تعریف‌شده در بخش پیشین، از الگوریتم بهینه‌سازی گرادیان مجاورتی (Proximal Gradient Algorithm)استفاده می‌شود .در این روش، بهینه‌سازی به صورت متناوب و تکرارشونده طی مراحل k (که $1 \leq k \leq K$) انجام می‌پذیرد .

الف) بهینه‌سازی نقشه روشنایی:(L_k)

در تکرار kام، تابع هدف نسبت به متغیر L به حداقل می‌رسد:

$$L_k = \arg\min_{L} \frac{1}{2}||I - R_{k-1} \odot L||_2^2 + \gamma\phi(L) + \frac{\lambda}{2}||w_k \odot \nabla L||_2^2$$

الگوریتم گرادیان مجاورتی این بهینه‌سازی را به دو مرحله تفکیک می‌کند: ترم نزول گرادیان (که یک متغیر کمکی $\hat{L}_k$ را تولید می‌کند) و ترم مجاورتی (Proximal term)که قیود را اعمال می‌کند :

$$\hat{L}_k = \frac{1}{2}||I - R_{k-1} \odot \hat{L}||_2^2 + \frac{\gamma}{2}||\hat{L} - L_{k-1}||_2^2 + \frac{\lambda}{2}||w_k \odot \nabla\hat{L}||_2^2$$

$$L_k = prox_{\phi}\left(R_{k-1}, \hat{L}_k\right)$$

با مشتق‌گیری مستقیم و برابر صفر قرار دادن آن، راه‌حل فرم بسته (Closed-form solution) برای متغیر $\hat{L}_k$ به دست می‌آید:

$$\hat{L}_k = (R_{k-1}^2 + \lambda w_k^2 \nabla^2 + \gamma\mathbf{1})^{-1}(R_{k-1}I + \gamma L_{k-1})$$

که در آن $\mathbf{1}$ یک ماتریس تماما یک (All-ones matrix) است و w_k بر اساس L_{k-1} محاسبه می‌گردد .

ب) بهینه‌سازی نقشه بازتاب:(R_k)

به طور مشابه، تابع هدف برای مؤلفه R با ثابت در نظر گرفتن L_k کمینه می‌شود:

$$R_k = \arg\min_{R} \frac{1}{2}||I - R \odot L_k||_2^2 + \beta\varphi(R) + \mu\mathcal{S}\left(\mathcal{A}(I) - \mathcal{A}(R)\right)$$

با تفکیک این معادله به ترم نزول گرادیان و ترم مجاورتی، راه‌حل فرم بسته برای پاسخ اولیه نقشه بازتاب ($\hat{R}_k$) به صورت زیر استخراج می‌گردد:

$$\hat{R}_k = (L_k^2 + \beta\mathbf{1} + Q_a)^{-1}(L_k I + \beta R_{k-1} + Q_b)$$

متغیرهای Q_a و Q_b شامل پارامترهای پیچیده‌ای از جمله ثابت لیپشیتز (Lipschitz constant) و عملگرهای نفوذ نمایی پرونا–مالیک هستند که در فاز بعد توسط معماری شبکه تقریب زده می‌شوند .

بخش چهارم: معماری شبکه و بسط معادلات(Deep Unfolding)

ایده اصلی شبکه‌های DUN این است که راه‌حل‌های تکرارشونده و فرمول‌های اثبات‌شده در بخش قبل، به لایه‌های یک شبکه عصبی چندمرحله‌ای به نام UnfoldIRبسط داده شوند .شبکه با تعداد مراحل پیش‌فرض $K = 3$ طراحی شده و هر مرحله شامل دو ماژول حیاتی است :

۱ .ماژول) RAIC اصلاح روشنایی با کمک بازتاب :(این ماژول وظیفه حل ترم مجاورتی روشنایی را بر عهده دارد .فرمول بسته محاسبه $\hat{L}_k$ به صورت یک لایه با پارامترهای قابل یادگیری پیاده‌سازی می‌شود .سپس برای جایگزینی عملگر مجاورتی $prox_{\phi}$، از یک فضای حالت بصری (Visual State Space - VSS) استفاده می‌گردد :

$$L_k = \mathcal{L}\left(\hat{L}_k, L_{k-1}\right) = VSS\left(\hat{L}_k, L_{k-1}\right)$$

استفاده از VSS به شبکه اجازه می‌دهد تا ویژگی‌های غیرمحلی (Non-local features) را استخراج کرده و نورپردازی را به صورت سراسری در تصویر تنظیم کند؛ این قابلیت در تصاویری که نور پس‌زمینه ناهمگون دارند) مانند (Backlit بسیار حیاتی است .

۲ .ماژول IGRE (بهبود بازتاب با هدایت روشنایی) و حلگر :ODE در حین اصلاح نقشه روشنایی، نویزهای پنهان در مناطق تاریک به شدت تقویت می‌شوند .با بسط معادله $I = R \odot L$ متوجه می‌شویم که برای رفع این نویزها بدون آسیب رساندن به بافت تصویر، نیاز به یک معماری آگاه به فرکانس (FVSS)است .فضای FVSS با استفاده از تبدیل موجک گسسته (DWT) تصویر را به باندهای فرکانسی تجزیه می‌کند :

$$R'_k = IDWT\left(VSS\left(DWT(\hat{R}_k)\right)\right)$$

سپس این خروجی، به همراه اطلاعات استخراج‌شده از نقشه روشنایی (L_k) مجددا وارد یک بلوک VSS می‌شود تا بافت‌های مشابه به صورت سراسری تراز شده و نقاط سالم تصویر، بازیابی نقاط تخریب‌شده را هدایت کنند .

نوآوری ریاضی در معماری :شبکه‌های پسماند (Residual) مرسوم، در واقع معادل گسسته‌سازی اویلر مرتبه اول از یک معادله دیفرانسیل معمولی (ODE)هستند که خطای برش قابل‌توجهی دارند .در معماریUnfoldIR ، ماژول FVSS درون یک چارچوب حلگر رانگ–کوتا مرتبه دوم (Runge-Kutta 2 - RK2) تعبیه شده است تا دقت معادلات عددی به شدت ارتقا یابد .فرمول نهایی بازتاب در این چارچوب عبارت است از :

$$R_k = \mathcal{R}(\hat{R}_k, L_k) = \hat{R}_k + g_w \tilde{R}_k + (1 - g_w)\tilde{R}_k^1$$

در این فرمول، g_w یک مکانیزم دروازه‌بندی (Gating mechanism) قابل یادگیری مبتنی بر عملگر سافت‌مکس (Softmax) است که انعطاف‌پذیری شبکه را در ادغام اطلاعات به حداکثر می‌رساند .

بخش پنجم: تابع هزینه سازگاری اطلاعات(ISIC Loss)

در شبکه‌های چندمرحله‌ای، ثبات در مراحل پایانی یک چالش بزرگ است. معماری UnfoldIR یک تابع هزینه اختصاصی به نام ISIC معرفی می‌کند که تنها در دو مرحله آخر شبکه اعمال می‌شود :

$$L_{ISIC} = ||R_k \odot L_{k-1} - R_k \odot L_k||_2 + ||\nabla(R_{k-1} \odot L_k) - \nabla(R_k \odot L_k)||_1$$

این تابع ریاضی به جای محدود کردن اجزای مستقل R) و L(، سازگاری را بر روی ترکیب آن‌ها (تصویر نهایی بازیابی‌شده) اعمال می‌کند .این رویکرد تضمین می‌کند که تغییرات جزئی روشنایی در مراحل پایانی، باعث تخریب جزئیات حیاتی بافت نشود .از آنجا که این تابع یک محدودیت درون‌شبکه‌ای و خودنظارتی (Self-supervised)است، به شبکه UnfoldIR اجازه می‌دهد تا در تنظیمات بدون ناظر (Unsupervised) نیز عملکرد پایداری از خود نشان دهد .

بخش ششم: ارزیابی کمی و دستاوردهای تجربی

برای اثبات برتری ریاضی و معماری شبکه پیشنهادی، آزمایش‌های گسترده‌ای انجام شده است:

- **ارزیابی در وظایف :IDIR** این مدل بر روی ۵ مجموعه داده مختلف از جمله LOL-v1/v2 (تصاویر کم‌نور) ، UIEB (تصاویر زیر آب)، BAID (تصاویر با نور پس‌زمینه) و تصاویر فوندوس پزشکی ارزیابی شده است .در تمامی این مجموعه‌ها، مدل UnfoldIR توانسته است در معیارهای کیفی PSNRو SSIM از برترین مدل‌های موجود) نظیر Reti-Diff و (CIDNet عبور کند و کمترین میزان انحراف معیار) مانند FID و (BIQE را ثبت کند .
- **ارتقای وظایف پایین‌دستی (Downstream Tasks)** : خروجی‌های قدرتمند این مدل، دقت شبکه‌های تشخیص شیء (مانند YOLO روی دیتاست ExDark) را افزایش داده و در تسک‌های پیچیده‌تری همچون سگمنتیشن معنایی و سگمنتیشن اشیای پنهان (Concealed Object) روی دادگان COD10K نتایج شگرفی به دست آورده است .
- **نسخه‌های سبک‌وزن :**معماری UnfoldIR علاوه بر مدل اصلی، در دو نسخه بسیار بهینه به نام‌های) UnfoldIR-t نسخه Tiny با تنها ۰.۰۹ میلیون پارامتر و ۰.۸۶ گیگافلاپس و UnfoldIR-s نسخه Small با ۰.۳۵ میلیون پارامتر (ارائه شده است .قدرت فرمول‌بندی‌های صریح ریاضی باعث شده تا حتی نسخه Tiny این شبکه، مدل‌های بسیار سنگین‌تر پیشین را شکست دهد که این دستاوردی کم‌نظیر در حوزه لبه‌ی تکنولوژی (Edge AI) است .