

## از فیلترهای کلاسیک تا شبکه‌های عصبی پیچشی: مبانی ریاضی و تکامل انتقال سبک در تصاویر

محمود امین طوسی<sup>۱</sup>

**چکیده.** انتقال سبک عصبی (Neural Style Transfer) یکی از نمونه‌های درخشان تلفیق ریاضیات، آمار، یادگیری عمیق و هنر است. در این مقاله، با نگاهی تاریخی و تحلیلی، سیر تحول ابزارهای ریاضی در پردازش تصویر از فیلترهای کلاسیک و مشتق‌پایه، تبدیلات فرکانسی و نمایش‌های چندمقیاسی تا شبکه‌های عصبی پیچشی که زیربنای فهم روش انتقال سبک عصبی هستند را دنبال می‌کنیم؛ و نشان داده می‌شود که چگونه این مفاهیم زمینه شکل‌گیری شبکه‌های عصبی پیچشی و در نهایت الگوریتم انتقال سبک را فراهم کرده‌اند. در ادامه، ساختار ریاضی انتقال سبک (استایل) با تمرکز بر نقش ماتریس گرام، فضاهای ویژگی یادگرفته‌شده و فرمول‌بندی تابع هدف بررسی می‌شود. نکته‌ی محوری در «انتقال سبک عصبی» تعریف تابع هدفی است که تنها در پرتو دانش شبکه‌های عصبی عمیق معنا می‌یابد؛ زیرا مفاهیم «سبک» و «محتوا» در فضای ویژگی‌های یادگرفته‌شده از داده‌ها بازنمایی می‌شوند، نه در فضای پیکسل‌ها. تعریف منبعث از مفاهیم آماری و حل این تابع هدف به کمک روش‌های کلاسیک بهینه‌سازی، نمود روشن پیوند میان ریاضیات کاربردی، آمار و یادگیری عمیق در یک خلاقیت هنری است.

### ۱. مقدمه

در دهه‌ی اخیر، ظهور یادگیری عمیق<sup>۱</sup> تحولی بنیادین در حوزه‌ی بینایی ماشین و پردازش تصویر ایجاد کرده است. یکی از جلوه‌های خاص این تحول، روشی موسوم به انتقال سبک عصبی<sup>۲</sup> است [۹]؛ الگوریتمی که با بهره‌گیری از شبکه‌های عصبی پیچشی (CNNs)<sup>۳</sup> می‌تواند سبک بصری یک تصویر را به تصویر دیگر منتقل سازد. شکل ۱ چند نمونه انتقال سبک عصبی را نشان می‌دهد. در ظاهر، این فرآیند نوعی ترکیب هنری از رنگ، بافت و فرم است، اما در عمق خود بر پایه‌ی مفاهیم ریاضی، آماری و محاسباتی استوار است؛ مفاهیمی همچون تجزیه‌ی سیگنال به مؤلفه‌های فرکانسی، فیلترهای مکانی و آماری، و تحلیل هم‌بستگی در فضای ویژگی‌ها.

پرسش بنیادینی که انتقال سبک عصبی به آن پاسخ داده است، این هست: چگونه می‌توان سبک یک تصویر را از محتوای آن جدا کرد و سپس با محتوای تصویری دیگر ترکیب نمود؟ پاسخ به این پرسش با ابزارهای کلاسیک پردازش تصویر میسر نبود، زیرا تعریف صوری «سبک» و «محتوا» نیازمند درکی از فضای ویژگی‌های چندلایه‌ای است که تنها با ظهور

رده‌بندی موضوعی ریاضیات ۲۰۲۰ (MSC 2020): کد، کد.

عبارات و کلمات کلیدی: انتقال سبک عصبی، شبکه‌های عصبی پیچشی، ماتریس گرام، تبدیلات فرکانسی، بازنمایی چندمقیاسی

نوع مقاله: ترویجی

دبیر تخصصی رابط: علیرضا امینی هرنیدی.

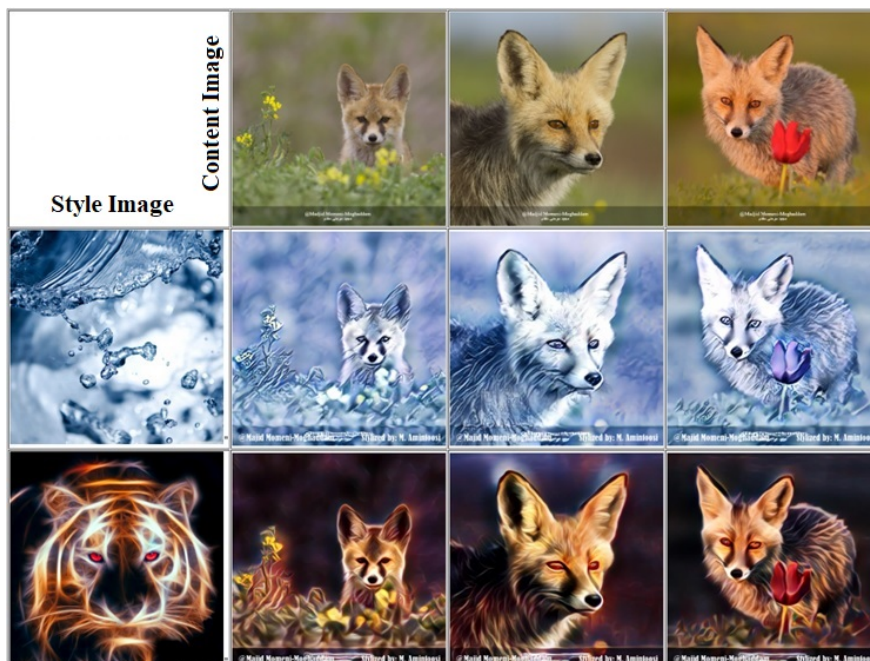
تاریخ دریافت: ۱۴۰۴/۰۸/۲۹ تاریخ پذیرش: ۱۴۰۵/۰۳/۲۴ تاریخ انتشار آنلاین: ۱۴۰۰/XX/XX

ارجاع به مقاله: م. امین طوسی، از فیلترهای کلاسیک تا شبکه‌های عصبی پیچشی: مبانی ریاضی و تکامل انتقال سبک در تصاویر، ریاضی و جامعه، -، (۱۴۰۰)، شماره -، صص. XX-XX.

<https://dx.doi.org/10.22108/msci.2026.147524.1771>

<sup>1</sup>deep learning <sup>2</sup>neural style transfer (NST) <sup>3</sup>convolutional neural networks

شبکه‌های عصبی پیچشی ممکن شد. در این چارچوب، ساختار ریاضی انتقال سبک بر پایه‌ی تابع هدفی<sup>۴</sup> تعریف می‌شود که دو مؤلفه‌ی اصلی دارد: تابع زیان محتوا و تابع زیان سبک. مؤلفه‌ی دوم، یعنی زیان سبک، بر اساس ماتریس گرام<sup>۵</sup> تعریف می‌شود که هم‌بستگی آماری بین ویژگی‌های لایه‌های مختلف شبکه را اندازه‌گیری می‌کند. این تابع هدف به‌گونه‌ای طراحی شده است که به حداقل رساندن آن از طریق روش‌های گرادیان‌کاهشی، موجب بازتولید تصویری می‌شود که محتوای تصویر اصلی را حفظ کرده ولی سبک تصویر مرجع را به خود گرفته است. این تلفیق زیبا از مفاهیم ریاضی، آمار و بهینه‌سازی با یادگیری عمیق، نمونه‌ای کم‌نظیر از کاربریست تئوری ریاضی، برنامه‌نویسی و تجربه در علوم کامپیوتر است.



شکل ۱. چند نمونه انتقال سبک عصبی، تصاویر آب و ببر، تصاویر سبک (استایل) و تصاویر روباه، تصاویر محتوا هستند. نتایج اجرای یک روش انتقال سبک عصبی [۲۰] در این شکل دیده می‌شود که استایل تصاویر آب و ببر بر روی تصاویر روباه اعمال شده است. (حق نشر تصاویر روباه، متعلق به مجید مؤمنی مقدم، عضو هیات علمی گروه زیست‌شناسی دانشگاه حکیم سبزواری است) نمونه تصاویر دیگر از آدرس <https://github.com/mamintoosi/MMM-Artistic-photos> قابل مشاهده هست.

**Figure 1:** Several examples of Neural Style Transfer: the water and tiger images are the *style* images, and the fox images are the *content* images. The results of executing a Neural Style Transfer method [20] are shown in this figure, where the style of the water and tiger images has been applied to the fox images. (Copyright of fox images belongs to Majid Momeni Moghaddam, faculty member of the Biology Department at Hakim Sabzevari University.) Additional sample images can be viewed at <https://github.com/mamintoosi/MMM-Artistic-photos>.

<sup>4</sup>objective function    <sup>5</sup>Gram matrix

هدف این مقاله آن است که با نگاهی تاریخی و تحلیلی، زمینه‌های ریاضی پیدایش «انتقال سبک عصبی» را روشن سازد. نوشتار حاضر از نوع «علمی-ترویجی» است و برای خواننده‌ای نگاشته شده که با مفاهیم پایه‌ای ریاضیات و آمار آشنایی دارد، اما لزوماً با یادگیری ماشین، پردازش تصویر و شبکه‌های عصبی آشنایی ندارد. بنابراین، تأکید اصلی بر نشان دادن پیوند میان دانش کلاسیک ریاضی و فناوری روز است، نه ارائه یک تحلیل پژوهشی بدیع یا پوشش جامع تمام روش‌های انتقال سبک. در این راستا، ابتدا ابزارهای کلاسیک نمایش و فیلترگذاری تصویر مرور می‌شوند، سپس مفاهیمی چون پیچش، تبدیل موجک و تحلیل مؤلفه‌های اصلی به عنوان ابزارهای سنتی نمایش داده‌ها و استخراج ویژگی مورد بررسی قرار می‌گیرند. در ادامه ساختار شبکه‌های پیچشی و نقش آن‌ها در نمایش سلسله‌مراتبی تصویر تبیین می‌شود، و نهایتاً الگوریتم انتقال سبک عصبی و برخی موارد مرتبط با آن بیان خواهد شد.

## ۲. مبانی ریاضی نمایش تصویر و فیلترهای کلاسیک

آشنایی با برخی مفاهیم بنیادی در ریاضیات کلاسیک سیگنال و تصویر، درک رفتار شبکه‌های عصبی پیچشی در پردازش تصویر، را تسهیل می‌کند. بسیاری از ایده‌های نوین در شبکه‌های عصبی پیچشی در واقع ادامه‌ی همان اصولی‌اند که دهه‌ها پیش در مهندسی برق و ریاضیات کاربردی توسعه یافته‌اند. در این بخش، به مرور مختصر این مبانی می‌پردازیم تا در بخش‌های بعد بتوانیم جایگاه ریاضی انتقال سبک عصبی را در این مسیر تحول درک کنیم.

### ۱.۲. نمایش ریاضی تصویر. تصویر دیجیتال را می‌توان به صورت یک تابع دوبعدی از مختصات مکان تعریف کرد:

$$(1) \quad f(x, y) : \Omega \subset \mathbb{Z}^2 \rightarrow \mathbb{R}^c$$

که در آن  $\Omega$  دامنه‌ی تصویر و  $c$  تعداد کانال‌های رنگی است. از دیدگاه ریاضی، هر تصویر را می‌توان ترکیبی خطی از توابع پایه دانست:

$$(2) \quad f(x, y) = \sum_n \alpha_n \phi_n(x, y)$$

که در آن  $\phi_n$ ها توابع پایه و  $\alpha_n$  ضرایب ترکیب‌اند. انتخاب این پایه‌ها، اساس تفاوت میان نمایش‌های مختلف تصویر است، از بسط‌های فوریه و موجک گرفته تا نمایش‌های داده‌محور در یادگیری عمیق.

۲.۲. فیلترهای کلاسیک و مفهوم فرکانس. فیلترگذاری<sup>۶</sup> نخستین گام در پردازش تصویر است و هدف آن، جداسازی اجزای ساختاری تصویر بر اساس میزان تغییرات فضایی (یا همان فرکانس) است. در ساده‌ترین تقسیم‌بندی، فیلترها به دو گروه عمده تقسیم می‌شوند:

فیلترهای پایین‌گذر.<sup>۷</sup> این فیلترها (صافی‌ها) تغییرات سریع در شدت روشنایی را حذف می‌کنند و برای مات‌سازی تصویر و کاهش نوفه (نویز) به کار می‌روند. از جمله معروف‌ترین این فیلترها، فیلتر گوسی<sup>۸</sup> است که با تابع زیر شناخته می‌شود:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

که در آن  $\sigma$  میزان هموارسازی را کنترل می‌کند. حاصل ضرب یا پیچش این هسته با تصویر، موجب حذف جزئیات ریز می‌شود.

<sup>۶</sup>filtering <sup>۷</sup>Low-Pass filters <sup>۸</sup>Gaussian filter

فیلترهای بالاگذر.<sup>۹</sup> برخلاف فیلترهای پایین‌گذر، این فیلترها بر جزئیات و تغییرات سریع شدت روشنایی در تصویر تمرکز دارند و به‌ویژه برای آشکارسازی مرزها و لبه‌ها به‌کار می‌روند. یک نمونه‌ی معروف آن عملگر لاپلاس است:

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2},$$

که در عمل با هسته‌های گسسته‌ای مانند ماتریس زیر پیاده‌سازی می‌شود:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

پیش از لاپلاس، عملگرهای مشتق‌پایه مانند سوبل<sup>۱۰</sup> و پریویت<sup>۱۱</sup> برای تقریب گرادیان‌های جزئی تصویر طراحی شده بودند. این فیلترها، مشتق در جهت‌های افقی و عمودی را محاسبه می‌کنند و از ترکیب آن‌ها بزرگی گرادیان به‌صورت زیر به‌دست می‌آید:

$$|\nabla f| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}.$$

کمیت  $|\nabla f|$  نشان‌دهنده شدت تغییرات مکانی در تصویر است و در نواحی لبه، که تغییر روشنایی ناگهانی وجود دارد، مقادیر بالایی دارد. به‌عبارت دیگر، نقشه گرادیان شاخصی از «مرزهای ساختاری» تصویر فراهم می‌کند.

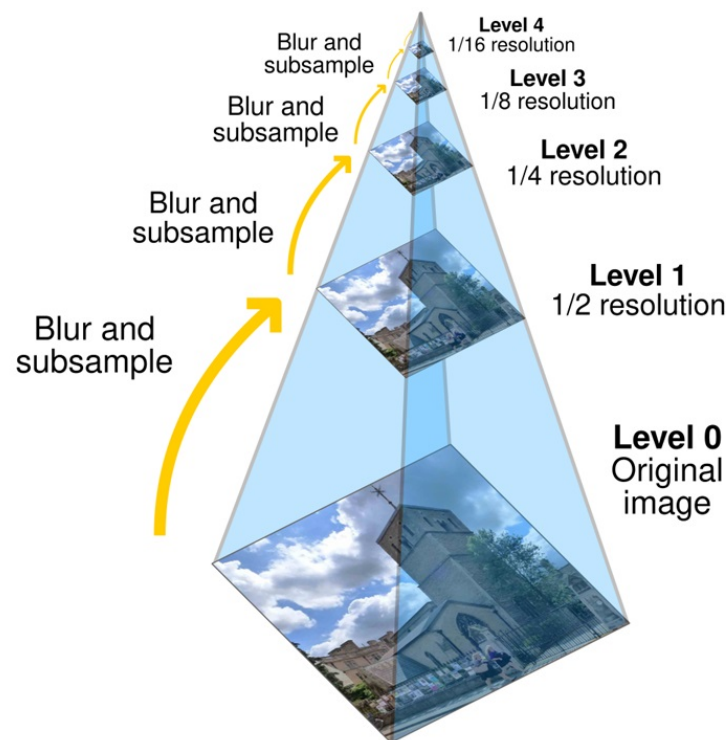
فیلترهای ترکیبی مانند LoG<sup>۱۲</sup> با هموارسازی تصویر توسط یک فیلتر گاوسی و سپس اعمال عملگر لاپلاس ساخته می‌شوند. این فیلتر در تشخیص نواحی با تغییرات شدید (لبه‌ها و نقاط برجسته) مؤثر است و با اعمال آن در مقیاس‌های مختلف، می‌توان ساختاری سلسله‌مراتبی از ویژگی‌های تصویر به‌دست آورد؛ در مقیاس‌های کوچک جزئیات ریز و در مقیاس‌های بزرگ‌تر ساختارهای کلی‌تر آشکار می‌شوند. شکل ۲ نمایی از پاسخ فیلتر LoG در سطوح مختلف مقیاس را نشان می‌دهد.<sup>۱۳</sup> در شبکه‌های عصبی پیچشی نیز ساختاری مشابه شکل می‌گیرد: در لایه‌های ابتدایی، فیلترهایی مشابه فیلترهای مشتق‌پایه و LoG به‌صورت خودکار آموخته می‌شوند، در حالی‌که لایه‌های عمیق‌تر بازنمایی‌های پیچیده‌تر و انتزاعی‌تر از داده را تشکیل می‌دهند. به این ترتیب، CNNها را می‌توان تداوم تکاملی همان ایده سلسله‌مراتبی در تحلیل تصویر دانست.

**۳.۲. نگاه فرکانسی و پایه‌های کلاسیک تا داده‌محور.** پس از فیلترهای مکانی و مشتق‌پایه، دیدگاه مکملی در ریاضیات و پردازش سیگنال شکل گرفت که تحلیل تصویر را در حوزه فرکانس ممکن می‌سازد. در این نگرش، تصویر به‌جای آن‌که صرفاً مجموعه‌ای از پیکسل‌ها باشد، ترکیبی از مؤلفه‌های نوسانی در مقیاس‌ها و جهت‌های مختلف در نظر گرفته می‌شود. تبدیل فوریه<sup>۱۴</sup>، تبدیل کسینوسی گسسته (DCT<sup>۱۵</sup>) و تبدیل موجک<sup>۱۶</sup> از مهم‌ترین ابزارهای این تحلیل‌اند. این تبدیلات مجموعه‌ای از پایه‌های متعامد فراهم می‌کنند که هر سیگنال یا تصویر را می‌توان با ترکیب خطی آن‌ها بازسازی کرد.

در تبدیل فوریه و DCT، هر پایه نمایانگر مؤلفه‌ای با فرکانس خاص است و انرژی تصویر عمدتاً در چند مؤلفه‌ی اول متمرکز می‌شود؛ ویژگی‌ای که مبنای فشرده‌سازی JPEG است. در مقابل، تبدیل موجک امکان تحلیل چندمقیاسی را فراهم می‌کند؛ بدین معنا که ساختارهای کلی در مقیاس‌های بزرگ و جزئیات ظریف در مقیاس‌های کوچک‌تر نمایان می‌شوند. همین ایده‌ی چندمقیاسی بعدها الهام‌بخش نسخه‌هایی از انتقال سبک عصبی شد که در آن، سبک در باندهای فرکانسی خاص اعمال می‌شود [۲۷].

<sup>۱۳</sup> تصویر برگرفته از این سایت است: [https://www.wikiwand.com/en/articles/Pyramid\\_\(image\\_processing\)](https://www.wikiwand.com/en/articles/Pyramid_(image_processing))

<sup>۹</sup>High-Pass filters <sup>۱۰</sup>Sobel <sup>۱۱</sup>Prewitt <sup>۱۲</sup>laplacian of gaussian <sup>۱۴</sup>Fourier ransform <sup>۱۵</sup>discrete cosine transform <sup>۱۶</sup>wavelet transform

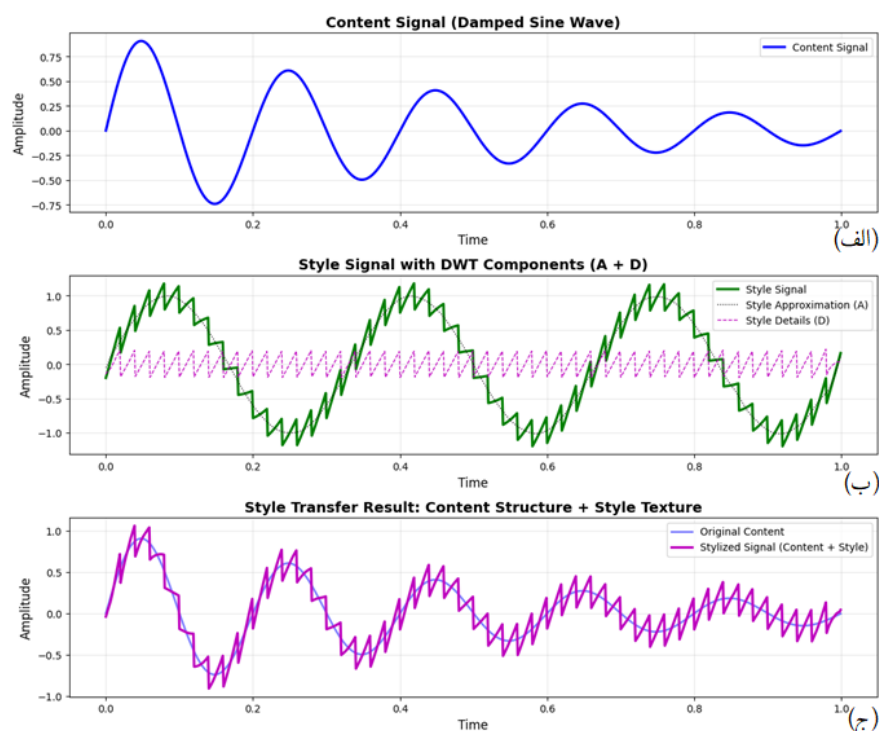


شکل ۲. نمایی شماتیک از پاسخ فیلتر LoG در مقیاس‌های مختلف و ساختار سلسله‌مراتبی حاصل از آن.

**Figure 2:** A schematic view of the Laplacian of Gaussian (LoG) filter response at different scales and the resulting hierarchical structure. (Image adapted from: [https://www.wikiwand.com/en/articles/Pyramid\\_\(image\\_processing\)](https://www.wikiwand.com/en/articles/Pyramid_(image_processing))).

با ظهور یادگیری عمیق، همین اصول ریاضی در قالبی داده‌محور باز تعبیر شدند. در شبکه‌های عصبی پیچشی، فیلترها دیگر از پیش تعیین شده نیستند، بلکه شبکه در فرایند آموزش، نگاشت‌هایی را می‌آموزد که به‌طور خودکار بیشترین اطلاعات مفید را از داده استخراج می‌کنند. هر لایه از شبکه را می‌توان مجموعه‌ای از نگاشت‌های غیرخطی دانست که داده را به فضای ویژگی جدیدی منتقل می‌کنند؛ فضایی که در آن، بازنمایی‌ها به‌صورت سلسله‌مراتبی سازمان می‌یابند؛ از لبه‌ها و بافت‌ها در لایه‌های اولیه تا ساختارهای مفهومی در لایه‌های عمیق‌تر. به این ترتیب، می‌توان سیر تاریخی تحلیل تصویر را از پایه‌های تحلیلی و متعامد (همچون فوریه و موجک) تا نگاشت‌های یادگرفته‌شده و داده‌محور (در شبکه‌های عصبی) پی گرفت. این گذار مفهومی، پیوندی میان ریاضیات کلاسیک، آمار و یادگیری عمیق برقرار می‌کند و بستر نظری لازم برای درک ریاضی الگوریتم‌های مدرنی همچون انتقال سبک عصبی را فراهم می‌سازد.

۴.۲. **مروری شهودی بر انتقال سبک مبتنی بر موجک.** برای درک شهودی مفهوم انتقال سبک عصبی، می‌توان از تبدیل موجک به‌عنوان ابزاری کلاسیک در تحلیل ساختار و بافت بهره گرفت. در این روش، سبک تصویری نه به‌صورت مستقیم روی پیکسل‌ها، بلکه بر ضرایب موجک اعمال می‌شود؛ بدین ترتیب کنترل دقیق‌تری بر میزان تأثیر سبک در جزئیات یا ساختار کلی تصویر به‌دست می‌آید.



شکل ۳. تجزیه موجک سیگنال‌ها و ترکیب سبک در یک بعد: (الف) سیگنال محتوا، (ب) سیگنال سبک به همراه مؤلفه‌های تقریب و جزئیات آن، (ج) ترکیب نهایی سیگنال محتوا و سبک دندان اره‌ای اخذ شده از سیگنال سبک.

**Figure 3:** Wavelet decomposition of signals and style combination in one dimension: (a) content signal, (b) style signal with its approximation and detail components, (c) final combination of the content signal and the sawtooth style extracted from the style signal.

در آزمایشی ساده، از تبدیل موجک برای اعمال سبک یک سیگنال یک بعدی برای سیگنالی دیگر بهره خواهیم گرفت. دو سیگنال را در نظر بگیرید: یکی سیگنال محتوا (مثلاً موج سینوسی میرا) که ساختار اصلی را دارد، و دیگری سیگنال سبک (سیگنالی همراه با نویز دندان اره‌ای) که حامل بافت و جزئیات است. با استفاده از تبدیل موجک گسسته<sup>۱۷</sup>، هر یک از این سیگنال‌ها به دو دسته مؤلفه تقسیم می‌شود: تقریب  $A$  و جزئیات  $D$ . سپس مؤلفه‌های تقریب از محتوای اصلی و جزئیات از سیگنال سبک گرفته می‌شود تا ترکیبی جدید ساخته شود (شکل ۳):

$$(۳) \quad f_{\text{stylized}}(t) = A_{\text{content}}(t) + D_{\text{style}}(t).$$

این آزمایش ساده، اگرچه صرفاً در سطح سیگنال انجام می‌شود، اما ایده‌ی کلیدی انتقال سبک عصبی را به خوبی نمایش می‌دهد: تفکیک و بازترکیب ساختار کلی (محتوا) و بافت‌های موضعی (سبک). در شبکه‌های عصبی پیچشی نیز همین ایده دنبال می‌شود؛ با این تفاوت که مؤلفه‌های موجک جای خود را به نگاشت‌های درونی لایه‌های شبکه می‌دهند: لایه‌های کم عمق تر

<sup>17</sup>discrete wavelet transform (DWT)



جدول ۱. مقایسه‌ی یادگیری ماشین کلاسیک و یادگیری عمیق.

Table 1: Comparison of classical machine learning and deep learning.

| ویژگی         | یادگیری ماشین کلاسیک    | یادگیری عمیق                    |
|---------------|-------------------------|---------------------------------|
| استخراج ویژگی | دستی (مانند PCA و HOG)  | خودکار (از طریق لایه‌های CNN)   |
| نوع پایه      | ثابت یا داده‌محور ساده  | داده‌محور پیچیده و غیرخطی       |
| ساختار مدل    | کم‌عمق و تفسیربردار     | عمیق و سلسله‌مراتبی             |
| محدوده‌ی عمل  | کل تصویر به‌صورت سراسری | نواحی محلی و ترکیب سلسله‌مراتبی |
| نیاز به داده  | نسبتاً کم               | بسیار زیاد                      |

نمایانگر جزئیات و سبک، و لایه‌های عمیق‌تر حامل ساختار و معنا هستند. از این شیوه‌ی کلی در [۱] برای افزایش وضوح تصویر و در [۲۷] برای انتقال سبک تصویر استفاده شده است.

**۵.۲. پایه‌های داده‌محور: از PCA تا شبکه‌های عصبی.** در قرن بیستم، ایده‌ای بنیادین شکل گرفت: به‌جای تعریف تحلیلی پایه‌ها، می‌توان آن‌ها را مستقیماً از داده آموخت. تحلیل مؤلفه‌های اصلی<sup>۱۸</sup> یکی از نخستین روش‌های یادگیری چنین پایه‌هایی است. در این روش، داده‌ها ابتدا از میانگین خود کم شده و سپس جهاتی در فضا انتخاب می‌گردند که بیشترین تغییرپذیری یا واریانس داده را در بر می‌گیرند. این جهت‌ها، که به آن‌ها مؤلفه‌های اصلی گفته می‌شود، همان محورهای هستند که ساختار آماری پنهان داده را بهتر از هر پایه‌ی از پیش تعریف شده‌ای آشکار می‌کنند. نتیجه‌ی این فرآیند، نمایش فشرده‌ای از داده است که در آن هر نمونه را می‌توان به‌صورت ترکیبی از میانگین و چند مؤلفه‌ی اصلی توصیف کرد که کاربردهای مختلفی در حوزه‌های مختلف مانند تشخیص چهره و تشخیص زمینه‌ی ویدئو دارد [۲]. برای نمونه، در تشخیص چهره، این مؤلفه‌ها به نام چهره‌های ویژه<sup>۱۹</sup> شناخته می‌شوند و هر تصویر چهره، ترکیبی خطی از چهره‌ی میانگین و تعدادی از این پایه‌های آموخته شده است. اهمیت این دیدگاه در آن است که مفهوم «یادگیری پایه‌ها از داده» بعدها در شبکه‌های عصبی عمیق بسط یافت؛ به‌طوری‌که لایه‌های اولیه‌ی شبکه‌ی پیچشی به‌صورت خودکار، پایه‌هایی مشابه مؤلفه‌های PCA را می‌آموزند، اما در سطوح بالاتر، نمایش‌هایی پیچیده‌تر و معنایی‌تر را استخراج می‌کنند. این پیوند مفهومی، گذار از روش‌های آماری کلاسیک مانند PCA به روش‌های یادگیری عمیق را نشان می‌دهد؛ گذری که زمینه‌ساز پیدایش روش‌های مدرن انتقال سبک عصبی شد و در حوزه انتقال سبک تصویر هم بکار برده شده است [۶].

**۶.۲. از یادگیری ماشین تا یادگیری عمیق.** در روش‌های کلاسیک یادگیری ماشین<sup>۲۰</sup>، ویژگی‌ها معمولاً به‌صورت دستی طراحی می‌شدند؛ برای نمونه با روش‌هایی چون PCA، فیلترهای لبه‌یاب، یا توصیف‌گرهای گرادیانی مانند<sup>۲۱</sup>. اما در یادگیری عمیق، خود شبکه به‌صورت خودکار و سلسله‌مراتبی ویژگی‌ها را از داده استخراج می‌کند و دیگر نیازی به طراحی دستی فیلترها نیست.

تفاوت مهم‌تر آن است که روش‌هایی مانند مویک یا PCA معمولاً بر کل تصویر به‌صورت یک واحد سراسری عمل می‌کنند، در حالی‌که شبکه‌های عصبی پیچشی روی نواحی محلی از تصویر، فیلتر اعمال می‌کنند و سپس با ترکیب لایه‌های پی‌درپی، ساختاری از ویژگی‌های ساده تا پیچیده را می‌سازند. این ویژگی موضعی، به شبکه‌ها قدرت بالایی در تشخیص الگوها و ترکیب جزئیات فضایی می‌دهد؛ توانایی که در روش‌های کلاسیک وجود نداشت. خلاصه تفاوت‌ها در جدول ۱ آمده است.

<sup>18</sup>principal component analysis (PCA) <sup>19</sup>eigenfaces <sup>20</sup>machine learning <sup>21</sup>histogram of oriented gradients (HOG)

در شبکه‌های عصبی پیچشی، هر لایه در واقع پایه‌ای جدید از نمایش داده را یاد می‌گیرد: لایه‌های اولیه لبه‌ها و رنگ‌ها را شناسایی می‌کنند، لایه‌های میانی بافت‌ها و ساختارهای محلی را، و لایه‌های عمیق‌تر مفاهیم سطح‌بالا را. از این منظر، شبکه‌های پیچشی را می‌توان بسط طبیعی ایده‌های پایه‌های داده‌محور دانست، اما با قدرت نمایش غیرخطی، سلسله‌مراتبی و مبتنی بر بهینه‌سازی. این دگرگونی بنیادی، زمینه‌ساز تعریف توابع هدفی چون تابع انتقال سبک شد، تابعی که تنها در فضای نمایش‌های درونی شبکه‌های عمیق معنا دارد.

### ۳. شبکه‌های عصبی پیچشی و استخراج ویژگی

همان‌گونه که در بخش قبل دیدیم، در روش‌های کلاسیک مانند PCA یا تبدیلات موجک، پایه‌ها به صورت خطی و سراسری تعریف می‌شوند و کل تصویر را به عنوان یک داده‌ی واحد در نظر می‌گیرند. این رویکرد گرچه برای تحلیل‌های کلی مفید است، اما نمی‌تواند الگوهای محلی و بافت‌های سبک‌های هنری را به خوبی بازنمایی کند. در مقابل، شبکه‌های عصبی پیچشی بر پایه‌ی مفهوم «میدان پذیرش محلی»<sup>۲۲</sup> عمل می‌کنند: هر فیلتر تنها بخش کوچکی از تصویر را می‌بیند و ویژگی‌های آن ناحیه را استخراج می‌کند. در لایه‌های بعدی، این ویژگی‌های محلی ترکیب می‌شوند تا ساختارهای بزرگ‌تر و انتزاعی‌تر ساخته شوند. به این ترتیب، شبکه به صورت سلسله‌مراتبی از لبه‌ها و رنگ‌ها تا اشیاء و مفاهیم سطح‌بالا را بازنمایی می‌کند. این توانایی در یادگیری نمایش‌های غیرخطی، محلی و چندلایه، زیربنای نظری روش‌های مدرن انتقال سبک عصبی است. در ادامه ابتدا مروری کوتاه بر اجزای اصلی شبکه‌های عصبی پیچشی داشته و سپس به چگونگی بازنمایی سبک (استایل) و بافت تصویر در این شبکه‌ها خواهیم پرداخت.

۱.۳. ساختار کلی شبکه‌های پیچشی. شبکه‌های عصبی پیچشی در دهه‌ی گذشته به ابزار اصلی در بینایی ماشین و پردازش تصویر بدل شدند. این شبکه‌ها از چند نوع لایه‌ی اصلی تشکیل شده‌اند: لایه‌های پیچشی برای استخراج ویژگی‌های محلی، توابع فعال‌سازی برای تبدیلات غیرخطی و لایه‌های تجمیعی برای کاهش ابعاد و افزایش پایداری. لایه پیچشی. در ساده‌ترین حالت، هر لایه پیچشی با اعمال یک فیلتر  $K$  بر تصویر ورودی  $I$  ویژگی‌های موضعی را استخراج می‌کند:

$$(4) \quad (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n) K(m, n).$$

در عمل، اغلب از همبستگی متقابل<sup>۲۳</sup> استفاده می‌شود، اما نقش هر دو مشابه است: فیلترها در سراسر تصویر حرکت می‌کنند و الگوهای محلی نظیر لبه‌ها و بافت‌ها را آشکار می‌سازند. مزیت مهم این ساختار، اشتراک وزن‌ها است که باعث می‌شود تعداد پارامترها مستقل از ابعاد ورودی باقی بماند.

تابع فعال‌سازی. پس از هر لایه پیچش، معمولاً تابعی غیرخطی مانند  $\text{ReLU}$ <sup>۲۴</sup> اعمال می‌شود:

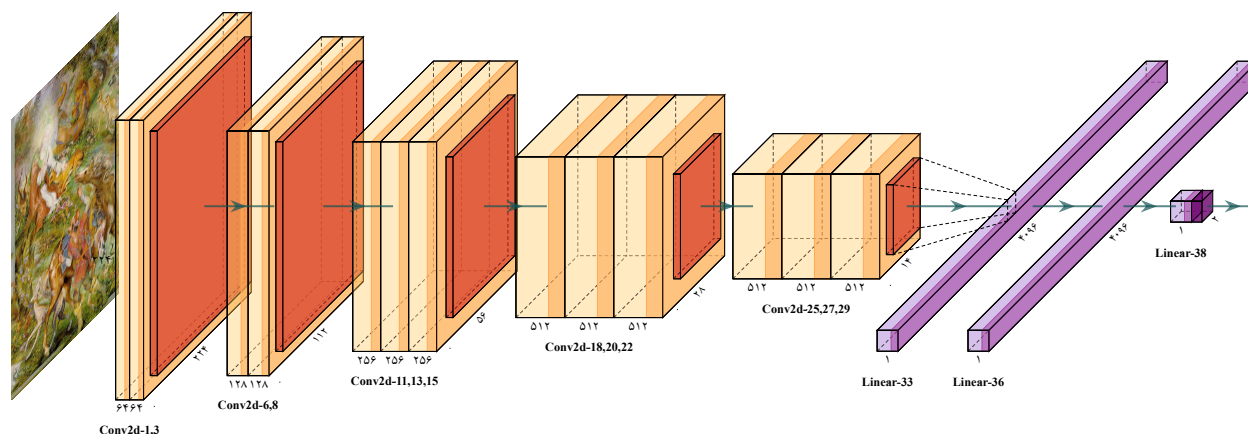
$$(5) \quad \text{ReLU}(x) = \max(0, x),$$

که منفی‌ها را حذف کرده و یادگیری روابط غیرخطی را ممکن می‌سازد. بدون چنین توابع غیرخطی، تمام شبکه صرفاً یک نگاشت خطی می‌بود.

لایه تجمیعی<sup>۲۵</sup>. برای کاهش ابعاد و افزایش مقاومت در برابر تغییرات مکانی، از لایه‌های تجمیعی استفاده می‌شود. در ادغام یا تجمیع بیشینه<sup>۲۶</sup> مقدار بیشینه در یک ناحیه انتخاب می‌شود، در حالی که در تجمیع میانگین<sup>۲۷</sup> میانگین مقادیر گرفته می‌شود. این عملیات موجب تمرکز تدریجی شبکه بر ساختارهای مهم‌تر می‌شود.

<sup>22</sup>local receptive field   <sup>23</sup>cross-correlation   <sup>24</sup>rectified linear unit   <sup>25</sup>pooling   <sup>26</sup>max pooling   <sup>27</sup>average pooling





شکل ۴. معماری مدل VGG16 [۲۳] با بیش از ۱۳۰ میلیون پارامتر (وزن). ورودی شبکه یک تصویر و خروجی آن برچسب طبقه متناظر با تصویر ورودی است. (بازنشر از [۳، شکل ۱])

**Figure 4:** Architecture of the VGG16 model [23] with over 130 million parameters (weights). The network input is an image and the output is the class label corresponding to the input image. (Image adapted from Figure 1 in [3]).

**۲.۳. یادگیری سلسله‌مراتبی ویژگی‌ها.** یکی از ویژگی‌های بنیادین شبکه‌های عصبی پیچشی، یادگیری سلسله‌مراتبی ویژگی‌هاست. در این شبکه‌ها، هر لایه در واقع پایه‌ای جدید از نمایش داده‌ها را می‌آموزد:

- لایه‌های ابتدایی، به الگوهای ساده‌ای مانند لبه‌ها و رنگ‌ها پاسخ می‌دهند؛
- لایه‌های میانی، این الگوها را به اشکال و بافت‌های ترکیبی تبدیل می‌کنند؛
- لایه‌های عمیق‌تر، ساختارهای معنایی سطح بالا مانند اشیاء و صحنه‌ها را بازنمایی می‌کنند.

از دیدگاه ریاضی، رفتار هر لایه را می‌توان با رابطه‌ی زیر توصیف کرد:

$$(۶) \quad f_l = \sigma_l(W_l * f_{l-1} + b_l),$$

که در آن  $f_l$  خروجی (ویژگی‌ها) در لایه‌ی  $l$ ،  $W_l$  وزن‌های قابل یادگیری،  $b_l$  بایاس، و  $\sigma_l$  تابع فعال‌سازی است. این فرایند، نوعی تجزیه‌ی چندسطحی از داده‌ها را ایجاد می‌کند که یادآور تبدیل‌های چندمقیاسی (مانند موجک و LoG) است، با این تفاوت که در شبکه‌های عصبی پیچشی، پایه‌ها از خود داده‌ها آموخته می‌شوند.

پیش از آن‌که به درک شهودی فرایند یادگیری ویژگی‌ها بپردازیم، مرور کوتاهی بر معماری‌های شبکه‌های عصبی-پیچشی، که برای ادامه کار ضروری است، خواهیم داشت. معماری‌های متداول این شبکه‌ها معمولاً شامل چندین لایه‌ی پیچشی، تابع فعال‌سازی و لایه‌های تجمیعی برای کاهش ابعاد هستند. در انتهای شبکه نیز معمولاً یک یا چند لایه‌ی پرسپترون چندلایه<sup>۲۸</sup> برای انجام طبقه‌بندی یا بازسازی قرار دارد. چنین ساختاری باعث می‌شود که شبکه بتواند نمایش‌هایی سلسله‌مراتبی و فزاینده از تصویر بسازد: از الگوهای ساده (لبه و رنگ) در لایه‌های اولیه تا مفاهیم پیچیده در لایه‌های انتهایی.

<sup>28</sup>multi-layer Perceptron (MLP)

داشتن مجموعه داده‌های بزرگ از الزامات اساسی آموزش این مدل‌ها است. این داده‌ها مستقیماً برای آموزش یا برای طراحی معماری‌های عمومی از پیش آموزش دیده به‌کار می‌روند. از مهم‌ترین این معماری‌ها می‌توان به مدل‌های VGG [۲۳]، ResNet [۱۱]، AlexNet [۱۸] و DarkNet [۲۲] اشاره کرد. هر یک از این شبکه‌ها از نظر عمق، نوع اتصال و ساختار لایه‌ها تفاوت‌هایی دارند، اما همگی بر اصل یادگیری تدریجی نمایش‌های درونی متکی هستند. اندازه‌ی این مدل‌ها معمولاً بسیار بزرگ است و شامل میلیون‌ها پارامتر می‌شود که در فرایند یادگیری تنظیم می‌شوند. شکل ۴ شمای کلی معماری VGG16 را نشان می‌دهد. خواننده‌ی علاقه‌مند برای آشنایی بیشتر با ساختار و کاربرد این شبکه‌ها می‌تواند به دو مقاله‌ی پیشین نویسنده مراجعه کند [۳، ۴].

**۳.۳. درک شهودی از یادگیری ویژگی‌ها.** مصورسازی فیلترها و ویژگی‌های لایه‌های مختلف، درک شهودی مهمی از نحوه‌ی عملکرد شبکه‌های عصبی پیچشی فراهم می‌کند. در لایه‌های اولیه، فیلترها اغلب مشابه فیلترهای کلاسیک پردازش تصویر (مانند فیلترهای سوبل یا گابور<sup>۲۹</sup>) هستند و به لبه‌ها، تغییرات شدت و بافت‌های ساده پاسخ می‌دهند. اما در لایه‌های میانی و عمیق‌تر، این فیلترها به تدریج الگوهای پیچیده‌تر و مفاهیم سطح بالا (مانند بخش‌های چهره، اجزای بدن، یا ساختار هندسی اشیاء) را بازنمایی می‌کنند.

یکی از پژوهش‌های تأثیرگذار در این زمینه، کار زیلر و فرگوس<sup>۳۰</sup> در مقاله‌ی [۲۸] است. آن‌ها با معرفی سازوکاری موسوم به DeconvNet، فعال‌سازی‌های درونی هر لایه را به فضای تصویر بازگردانند و نشان دادند که هر لایه از شبکه چه نوع ویژگی‌هایی را «می‌بیند». به کمک این روش، مشخص شد که ساختار سلسله‌مراتبی شبکه - از لبه‌ها تا مفاهیم پیچیده - در واقع نوعی سازمان‌دهی تدریجی ویژگی‌ها است. این درک شهودی در تبیین جداسازی محتوا و سبک در شبکه‌های انتقال سبک عصبی نیز نقش مهمی ایفا می‌کند. در شکل ۵ نمونه‌ای از چنین بازنمایی‌هایی نمایش داده شده است: فعال‌سازی‌های متناظر در لایه‌های مختلف شبکه شان می‌دهند چگونه فیلترها به تدریج از اجزای ساده به ساختارهای معنایی می‌رسند. در ادامه خواهیم دید که این ساختار سلسله‌مراتبی یادگیری ویژگی‌ها، مبنای تعریف ریاضی مفهوم «محتوا» و «سبک» در شبکه‌های عمیق است.

**۴.۳. نمایش تصویر در فضای ویژگی‌های شبکه‌های عصبی پیچشی.** از دیدگاه ریاضی، می‌توان یک شبکه‌ی عصبی پیچشی را به صورت دنباله‌ای از نگاشت‌ها در نظر گرفت:

$$(۷) \quad \Phi = \Phi_L \circ \Phi_{L-1} \circ \dots \circ \Phi_1,$$

که در آن  $\Phi_l$  عمل نگاشت از خروجی لایه‌ی  $l-1$  به  $l$  است. ویژگی‌های استخراج‌شده از تصویر  $I$  در لایه‌ی  $l$  به صورت زیر نوشته می‌شود:

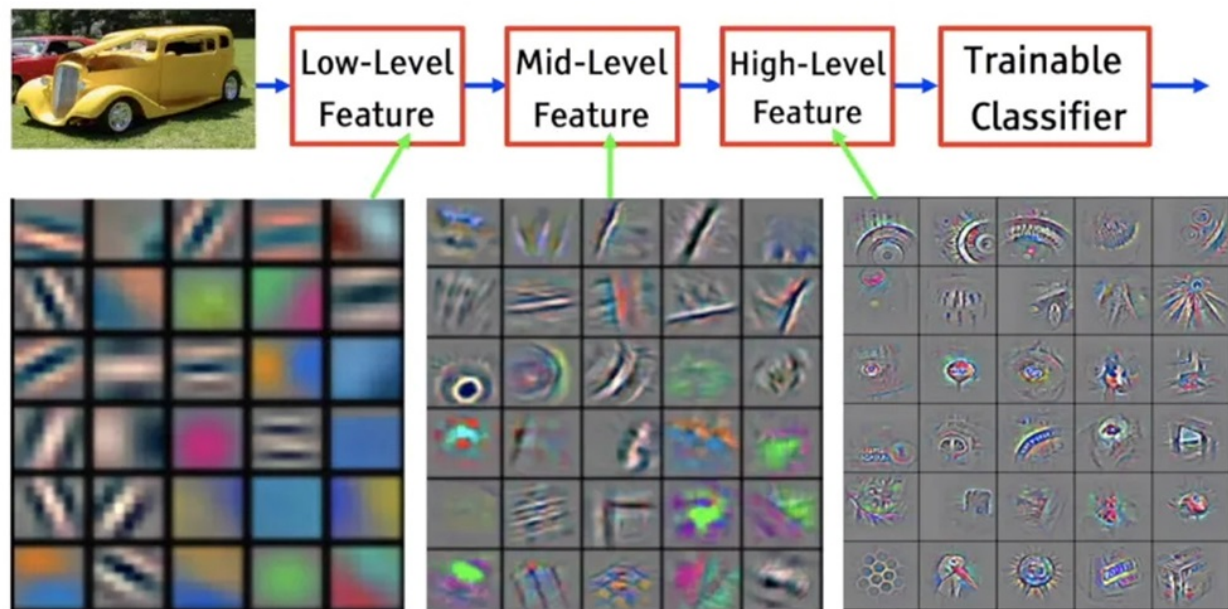
$$(۸) \quad F^l = \Phi_l(I).$$

این ویژگی‌ها به صورت سلسله‌مراتبی از بازنمایی‌های محلی و بافتی در لایه‌های ابتدایی، به بازنمایی‌های معنایی و مفهومی در لایه‌های عمیق‌تر تبدیل می‌شوند. برای سنجش شباهت محتوایی میان دو تصویر  $I_1$  و  $I_2$ ، می‌توان فاصله‌ی اقلیدسی میان نمایش‌هایشان در لایه‌ای خاص ( $l_c$ ) را اندازه‌گیری کرد:

$$(۹) \quad d_{l_c}(I_1, I_2) = \|\Phi_{l_c}(I_1) - \Phi_{l_c}(I_2)\|_2.$$

این معیار، اساس تعریف «شباهت محتوایی» در انتقال سبک عصبی است.

<sup>29</sup>Gabor <sup>30</sup>Zeiler & Fergus



شکل ۵. نمونه‌ای از واکنش فیلترهای مختلف در شبکه‌های عصبی پیچشی نسبت به یک تصویر ورودی. در لایه‌های ابتدایی، ویژگی‌های ساده مانند لبه‌ها آشکار می‌شوند؛ در حالی که لایه‌های عمیق‌تر به ترکیب‌های پیچیده‌تر و مفاهیم سطح بالا حساس‌اند [۲۵].

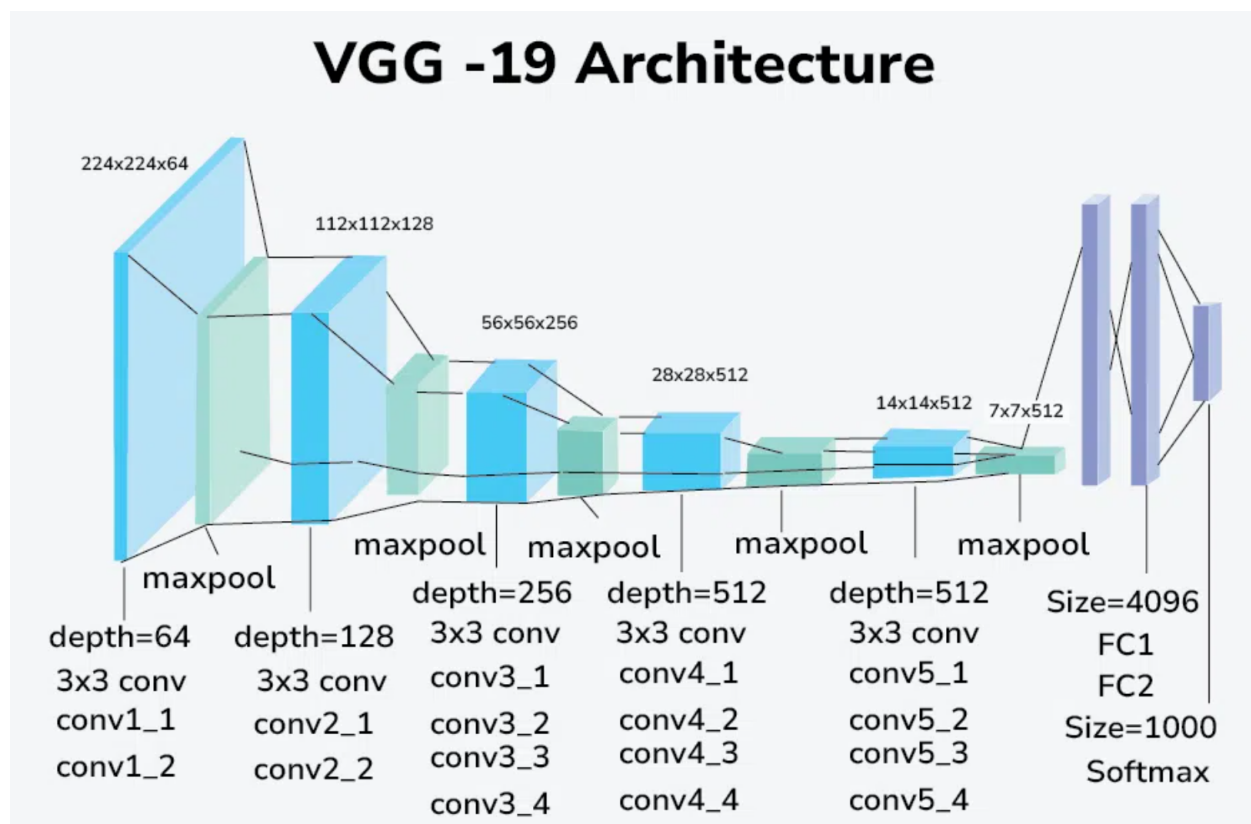
**Figure 5:** A sample of the response of different filters in Convolutional Neural Networks to an input image. In early layers, simple features such as edges are revealed; while deeper layers are sensitive to more complex combinations and high-level concepts [25].

در عمل، الگوریتم‌های انتقال سبک از شبکه‌هایی بهره می‌برند که پیش‌تر بر روی مجموعه داده‌های بزرگ مانند ImageNet آموزش دیده‌اند. از معماری‌های متداول می‌توان به مدل‌های ۱۶ لایه و ۱۹ لایه VGG اشاره کرد که ساختار کلی مدل ۱۶ لایه در شکل ۴ نشان داده شد. شکل ۶ نمایشی از یک شبکه‌ی VGG-19 است که شامل ۱۹ لایه‌ی قابل یادگیری است (۱۶ لایه‌ی پیچشی و سه لایه‌ی طبقه‌بندی).

مطابق تحلیل گاتیس و همکاران [۹]، لایه‌های مختلف شبکه‌ی VGG بازنمایی‌هایی با ماهیت متفاوت از تصویر ارائه می‌دهند. در مراحل ابتدایی شبکه (مانند conv1\_1 و conv2\_1 در شکل ۶)، پاسخ فیلترها به تغییرات محلی و بافت‌های ظریف بسیار حساس است؛ در حالی که در لایه‌های عمیق‌تر (مانند conv4\_2 و conv5\_2)، جزئیات پیکسلی از میان می‌رود و تنها ساختار کلی و چیدمان اجسام حفظ می‌شود. به همین دلیل، ویژگی‌های لایه‌های بالاتر نماینده‌ی «محتوا» و لایه‌های پایین‌تر نماینده‌ی «سبک» در نظر گرفته می‌شوند.

در پیاده‌سازی اصلی انتقال سبک عصبی، گاتیس و همکاران برای نمایش محتوا از لایه‌ی conv4\_2 و برای نمایش سبک از لایه‌های conv1\_1، conv2\_1، conv3\_1، conv4\_1، conv5\_1 استفاده کردند [۹، ۱۴].

ترکیب چندین لایه‌ی سبک، باعث شد تا شبکه بتواند آماره‌های بافت را در مقیاس‌های گوناگون (از جزئیات ظریف تا ساختار کلی) منطبق کند و در نتیجه، جلوه‌ای هموارتر و طبیعی‌تر از انتقال سبک به‌دست آید. در مقابل، اگر لایه‌ی محتوا از



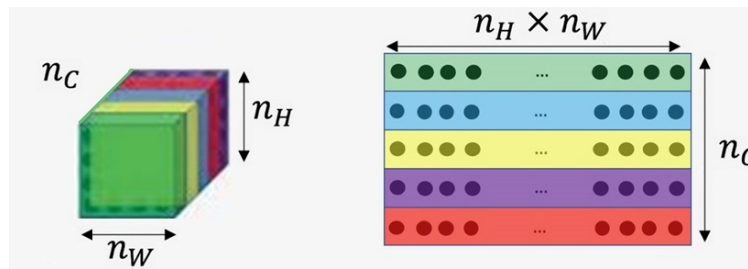
شکل ۶. نمایشی از معماری مدل VGG19؛ بازنشر از [۲۹، شکل ۸].

Figure 6: Illustration of the VGG19 model architecture; reproduced from [29, Figure 8].

سطوح پایین‌تر انتخاب شود، جزئیات ناخواسته‌ای مانند لبه‌ها و رنگ‌های اولیه نیز در تصویر نهایی باقی می‌مانند؛ در حالی که تطبیق محتوا در لایه‌ای بالاتر امکان دگرگونی سبک در عین حفظ مفهوم اصلی صحنه را فراهم می‌کند. به این ترتیب، ساختار سلسله‌مراتبی شبکه نه تنها ابزار استخراج ویژگی، بلکه بستری برای جداسازی مؤلفه‌های «محتوا» و «سبک» است مبنایی که هسته‌ی ریاضی الگوریتم گاتیس را تشکیل می‌دهد.

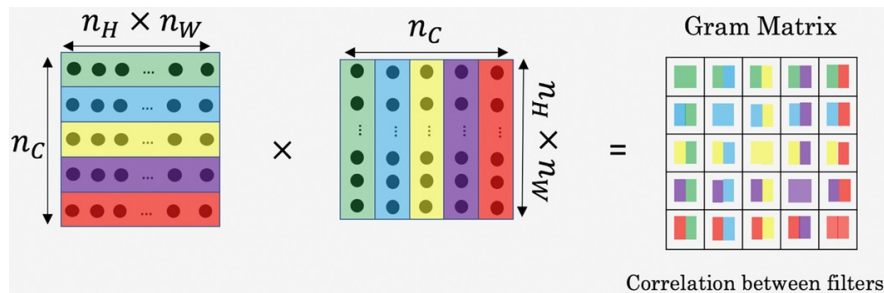
#### ۴. تعریف ریاضی و آماری سبک در تصاویر

در بخش‌های پیشین دیدیم که چگونه ابزارهای کلاسیک پردازش تصویر، از فیلترهای پایین‌گذر و بالاگذر تا تبدیلات موجک و فوریه، درک ما را از ساختار و الگوهای بصری شکل داده‌اند. اکنون می‌خواهیم ببینیم که در چارچوب شبکه‌های عصبی عمیق، مفهوم «سبک» چگونه تعریف و به صورت ریاضی مدل می‌شود. در واقع، انتقال سبک عصبی پلی میان ریاضیات آماری و یادگیری عمیق می‌سازد و اجازه می‌دهد که «ویژگی‌های آماری» یک تصویر به عنوان سبک آن شناخته شوند. درک مفهوم «سبک» در چارچوب شبکه‌های عصبی پیچشی (CNN) نیازمند آن است که ابتدا نحوه‌ی بازنمایی درونی تصاویر در این شبکه‌ها را بشناسیم. همان‌گونه که ذکر شد مطالعات زیلر و فرگوس [۲۸] نشان دادند که فیلترهای لایه‌های مختلف در CNN، الگوهایی از ساده‌ترین ساختارها (لبه‌ها و بافت‌های محلی) تا مفاهیم انتزاعی‌تر را در سطوح بالاتر بازنمایی می‌کنند.



شکل ۷. تبدیل یک تانسور ویژگی سه‌بعدی  $(C \times H \times W)$  به ماتریس دوبعدی  $C \times M$ . هر ردیف پاسخ یک فیلتر را در تمام مکان‌های فضایی نشان می‌دهد.

**Figure 7:** Converting a three-dimensional feature tensor  $(C \times H \times W)$  to a two-dimensional matrix  $C \times M$ . Each row shows the response of one filter at all spatial locations.



شکل ۸. تشکیل ماتریس گرام از نقشه‌های ویژگی: حاصل ضرب ماتریس  $F$  در ترانپوز آن در ترانپوز خود روابط آماری بین فیلترها را به دست می‌دهد.

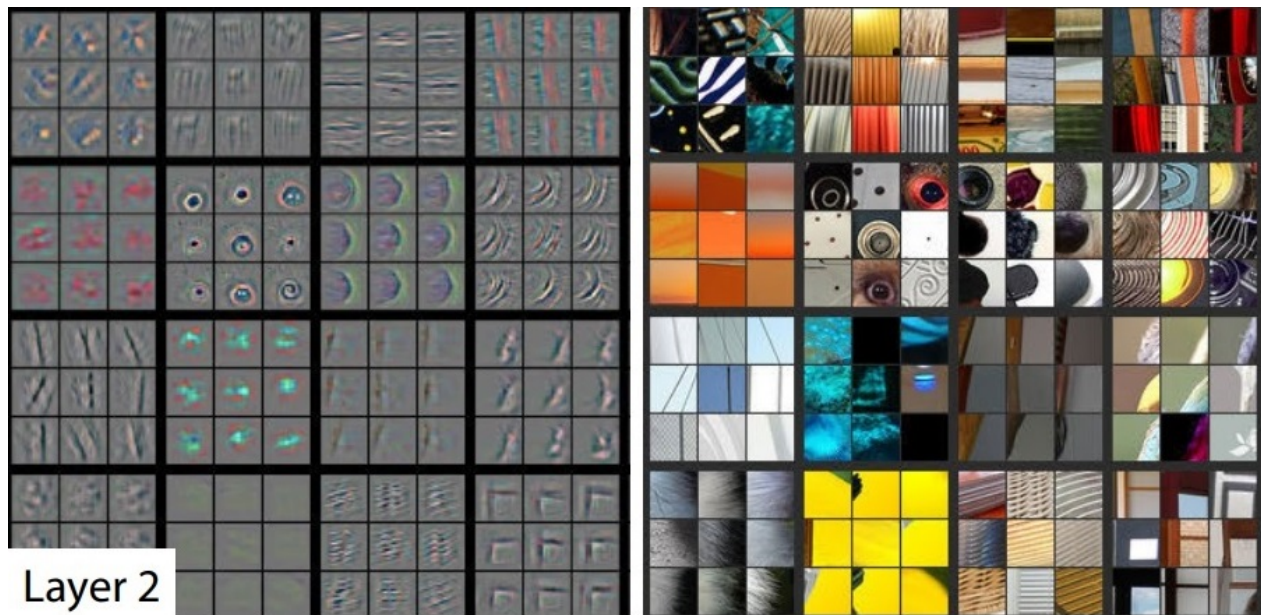
**Figure 8:** Formation of the Gram matrix from feature maps: the product of matrix  $F$  with its transpose yields the statistical relationships between filters.

در همین چارچوب، گاتیس و همکاران [۹] دریافتند که روابط آماری میان این فعال‌سازی‌ها، جوهری آن چیزی است که به عنوان «سبک» یک تصویر درک می‌کنیم.

۱.۴. از نقشه‌ی ویژگی تا بازنمایی سبک. فرض کنید یک تصویر ورودی  $I$  از میان یک شبکه‌ی CNN عبور داده می‌شود و در یکی از لایه‌های میانی، خروجی یا نقشه‌ی ویژگی آن به صورت  $F \in \mathbb{R}^{C \times H \times W}$  حاصل می‌گردد. در این جا  $C$  تعداد کانال‌ها (یا فیلترها)، و  $H$  و  $W$  به ترتیب ارتفاع و عرض نقشه‌ی ویژگی هستند. برای سهولت محاسبات، این تانسور سه‌بعدی را می‌توان به ماتریسی دوبعدی  $F \in \mathbb{R}^{C \times M}$  با  $M = H \times W$  بازآرایی کرد؛ یعنی هر سطر نماینده‌ی پاسخ‌های یک فیلتر در تمام موقعیت‌های مکانی است (شکل ۷).

۲.۴. ماتریس گرام: زبان ریاضی سبک. گاتیس و همکاران [۹] برای نخستین بار نشان دادند که همبستگی‌های آماری بین پاسخ فیلترها، می‌تواند ساختار سبک را در یک تصویر به خوبی بازنمایی کند. این همبستگی‌ها در قالب ماتریس گرام





شکل ۹. تجسم فیلترهای شبکه‌ی عصبی پیچشی در سطوح مختلف (بازنشر از [۲۸، شکل ۲]). همبستگی بین فیلترهایی که به ویژگی‌های گوناگون - مثلاً بافت، رنگ یا شکل - پاسخ می‌دهند، می‌تواند بیانگر ساختارهای ترکیبی‌تر (مانند پوست حیوانات یا الگوهای تزئینی) باشد.

**Figure 9:** Visualization of Convolutional Neural Network filters at different levels (reproduced from [28, Figure 2]). The correlation between filters responding to various features—for example, texture, color, or shape—can indicate more composite structures (such as animal skins or decorative patterns).

تعریف می‌شوند. برای ماتریس ویژگی  $F \in \mathbb{R}^{C \times M}$ ، ماتریس گرام  $G \in \mathbb{R}^{C \times C}$  به شکل زیر تعریف می‌شود:

$$(۱۰) \quad G = FF^T, \quad G_{ij} = \sum_{k=1}^M F_{ik}F_{jk}.$$

هر درایه‌ی  $G_{ij}$  میزان همبستگی میان فیلترهای  $i$  و  $j$  را بیان می‌کند. درایه‌های قطری  $G_{ii}$  قدرت پاسخ هر فیلتر و عناصر خارج از قطر میزان هم‌رخدادی یا هم‌تغییری فیلترها را منعکس می‌کنند (شکل ۸). از دیدگاه آماری، ماتریس گرام، همبستگی بین پاسخ‌های فیلترهای مختلف را در یک لایه نشان می‌دهد. اگر پاسخ فیلتر  $i$  در مکان‌هایی از تصویر که فیلتر  $j$  نیز فعال است، مقدار بالایی داشته باشد، درایه‌ی متناظر  $G_{ij}$  بزرگ خواهد بود؛ یعنی این دو فیلتر اغلب به صورت هم‌زمان برانگیخته می‌شوند. این هم‌رخدادی می‌تواند نشان‌دهنده‌ی وجود یک الگوی ترکیبی در تصویر باشد. برای مثال، فرض کنید یکی از فیلترها به خطوط راه‌راه حساس باشد و فیلتر دیگری رنگ نارنجی را تشخیص دهد. اگر در یک تصویر این دو ویژگی به طور هم‌زمان و در نواحی مشابه ظاهر شوند، مقدار  $G_{ij}$  میان این دو فیلتر افزایش می‌یابد؛ و بیانگر آن است که وقوع هم‌زمان خطوط راه‌راه و رنگ نارنجی نوعی بافت یا طرح خاص (مانند پوست ببر) را مشخص می‌کند (شکل ۹).



این بینش، توضیح شهودی روش انتقال سبک گاتیس و همکاران [۹] را تشکیل می‌دهد: سبک یک تصویر را می‌توان با روابط آماری بین فیلترهایی که به رنگ‌ها، بافت‌ها و ساختارهای مختلف پاسخ می‌دهند، توصیف کرد. به عبارت دیگر، سبک نه در مقادیر مکانی دقیق پیکسل‌ها، بلکه در الگوهای با هم فعال‌شده فیلترها نهفته است. همان‌گونه که زیلر و فرگوس [۲۸] نشان دادند، فیلترهای سطوح مختلف شبکه‌ی پیچشی از لبه‌ها و رنگ‌ها در لایه‌های اولیه تا مفاهیم ترکیبی‌تر در لایه‌های عمیق‌تر را بازنمایی می‌کنند. بنابراین، محاسبه‌ی ماتریس گرام از چندین لایه، نمایی چندمقیاسه از هم‌بستگی‌های آماری میان این مفاهیم را فراهم می‌سازد. مطابق یافته‌های زیلر و فرگوس [۲۸]، هرچه به لایه‌های عمیق‌تر در شبکه نزدیک شویم، فیلترها مفاهیم انتزاعی‌تری را بازنمایی می‌کنند (شکل ۵). به همین دلیل، محاسبه‌ی ماتریس گرام از چندین لایه‌ی مختلف باعث می‌شود سبک نهایی شامل اطلاعات آماری از سطوح گوناگون - از بافت‌های ظریف تا الگوهای کلی - باشد.

به صورت خلاصه، می‌توان گفت، تعریف ریاضی سبک در شبکه‌های پیچشی ترکیبی از این روابط آماری در لایه‌های مختلف است، که به صورت بردار یا مجموعه‌ای از ماتریس‌های گرام نمایش داده می‌شود. این بازنمایی، پلی میان ریاضیات آماری و ادراک بصری شبکه‌های عصبی است و مبنای روش‌های مدرن انتقال سبک عصبی را تشکیل می‌دهد.

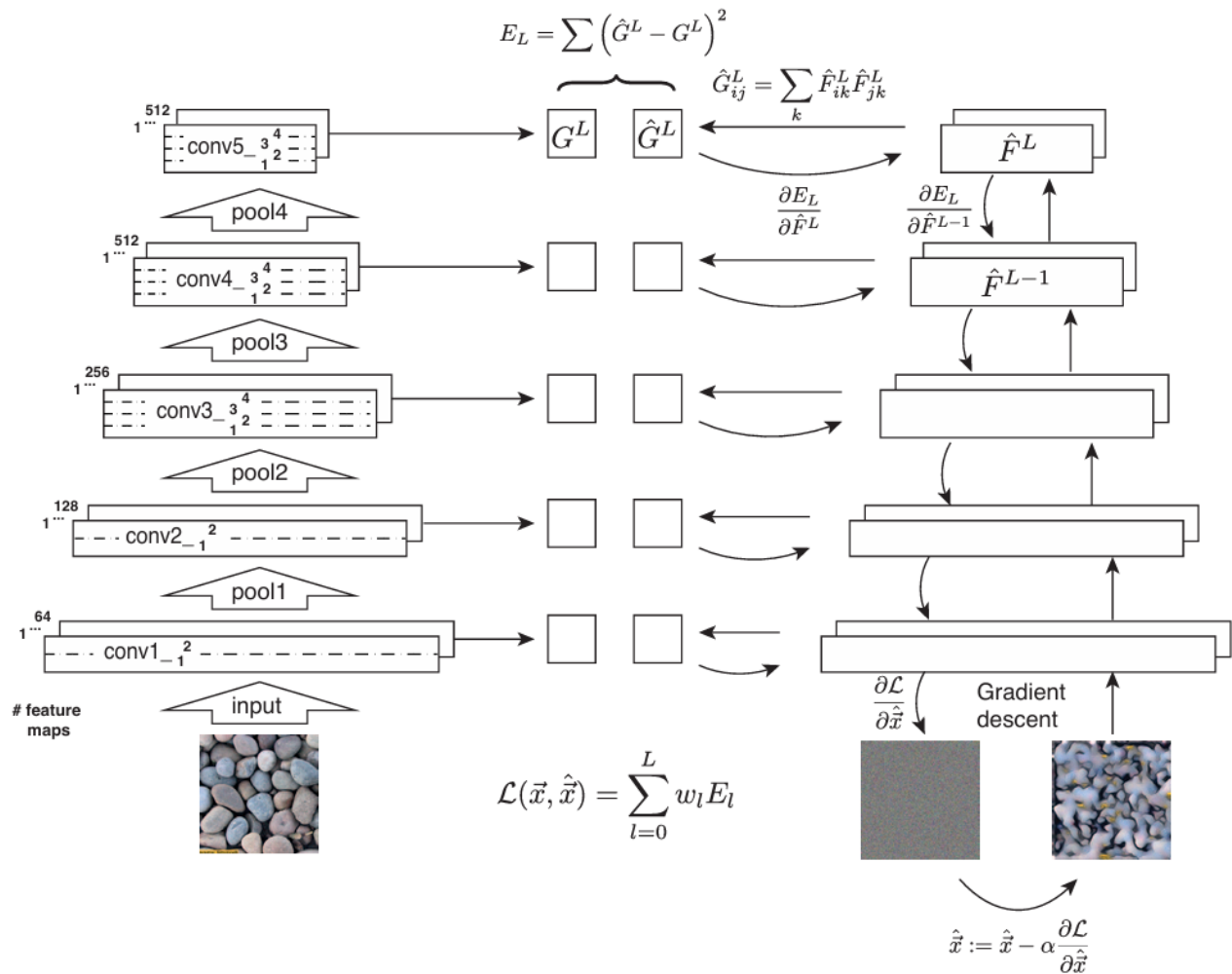
**۳.۴. تولید بافت‌های جدید با ماتریس گرام.** یکی از نخستین و زیباترین کاربردهای ایده‌ی ماتریس گرام در شبکه‌های عصبی پیچشی، تولید بافت‌های مصنوعی است [۱۰]. در این روش، هدف آن است که تنها با در اختیار داشتن یک تصویر نمونه از بافت، تصویری جدید بسازیم که از نظر آماری و ادراکی با آن هم‌خوانی داشته باشد. مثلاً اگر تصویر اصلی، یک تصویر از سنگ‌ریزه‌ها باشد، تصاویر مشابه با آن تولید شود. یک کاربرد آن می‌تواند تولید الگوها و بافت‌هایی مشابه طرح‌های کاشی‌ها و کاغذدیواری‌ها باشد که بر اساس یک نمونه تصویر انجام شود.

ابتدا تصویر نمونه‌ی بافت  $I_{\text{texture}}$  از چندین لایه‌ی شبکه‌ی CNN عبور داده می‌شود و ماتریس‌های گرام  $\{G_l^{\text{target}}\}$  مربوط به هر لایه محاسبه می‌گردد. این ماتریس‌ها هم‌بستگی میان فیلترهای فعال در هر لایه را ثبت می‌کنند و بدین ترتیب «سبک» آماری تصویر را در مقیاس‌های مختلف به صورت فشرده بازنمایی می‌کنند. در مرحله‌ی بعد، یک تصویر اولیه‌ی تصادفی (مثلاً نویز سفید) به عنوان ورودی انتخاب می‌شود و با عبور مکرر از شبکه و به روز رسانی مقادیر پیکسل‌ها از طریق گرادیان نزولی، به تدریج به تصویری تبدیل می‌شود که ماتریس‌های گرام آن با مقادیر هدف برابر باشند. به بیان دیگر، تابع زیان کلی به صورت ترکیبی از اختلاف گرام‌ها در چندین لایه تعریف می‌شود:

$$(11) \quad \mathcal{L}(\vec{x}, \hat{\vec{x}}) = \sum_l w_l \|G_l^{\text{current}} - G_l^{\text{target}}\|_F^2$$

که در آن  $w_l$  وزن هر لایه و  $\|\cdot\|_F$  نرم فروبنیوس است. این فرآیند در شکل ۱۰ (برگرفته از مقاله‌ی گاتیس و همکاران [۱۰]) نمایش داده شده است. در سمت چپ، تحلیل بافت با عبور تصویر اصلی از شبکه و محاسبه‌ی ماتریس‌های گرام دیده می‌شود، و در سمت راست، روند بهینه‌سازی تصویر نویزی برای بازتولید همان آمار بافت نمایش داده شده است. این الگوریتم نشان می‌دهد که تنها با استفاده از هم‌بستگی میان ویژگی‌ها می‌توان بافت‌های بصری پیچیده و طبیعی را بازتولید کرد.

با وجود نقش بنیادین ماتریس گرام در بازنمایی سبک، پژوهش‌های بعدی نشان دادند که این روش در برخی موارد با چالش‌هایی همراه است (برای نمونه، نادیده گرفتن آرایش فضایی ویژگی‌ها و بازتولید بافت‌های تکراری [۱۴]). به همین دلیل، روش‌های جایگزینی مانند نرمال‌سازی تطبیقی [۱۲] و رویکردهای مبتنی بر تبدیلات خطی [۶] پیشنهاد شده‌اند که هر کدام تعادلی میان کیفیت، سرعت و حفظ محتوا برقرار می‌کنند.



شکل ۱۰. نمای شماتیک از روش تحلیل و تولید بافت در مقاله‌ی [۱۰]. در سمت چپ، بافت ورودی از شبکه عبور داده می‌شود و ماتریس‌های گرام در سطوح مختلف محاسبه می‌گردد؛ در سمت راست، تصویر نویزی اولیه با گرادینان نزولی به‌گونه‌ای بهینه می‌شود که همان ماتریس‌های گرام را بازتولید کند و تصویر نویزی به بافتی مشابه با تصویر ورودی تبدیل شود.

**Figure 10:** Schematic view of the texture analysis and generation method from [10]. On the left, the input texture is passed through the network and Gram matrices are computed at different levels; on the right, the initial noise image is optimized via gradient descent to reproduce the same Gram matrices, transforming the noise into a texture similar to the input image.

ایده‌ی تولید بافت با ماتریس گرام نقطه‌ی آغاز درک آماری سبک در شبکه‌های عصبی بود و گاتیس و همکاران در پژوهش بعدی خود [۹] نشان دادند می‌توان از همین ایده‌ی آماری برای آفرینش تصاویری بهره گرفت که نه تنها سبک یک تصویر را بازنمایی می‌کنند، بلکه ساختار و محتوای تصویر دیگری را نیز حفظ می‌نمایند که همان «انتقال سبک عصبی» است.

## ۵. از تولید بافت تا انتقال سبک عصبی

پس از درک مفاهیم آماری نهفته در ماتریس گرام، اکنون به سراغ مسئله اصلی می‌رویم: چگونه می‌توان محتوای یک تصویر را با سبک تصویری دیگر ترکیب کرد؟، الگوریتمی که در سال ۲۰۱۶ توسط گاتیس و همکاران [۹] معرفی شد، کاربردی جذاب از آمار، بهینه‌سازی ریاضی و شبکه‌های عصبی پیچشی بود. در این روش، هیچ مدلی آموزش داده نمی‌شود؛ بلکه خود پیکسل‌های تصویر به صورت مستقیم بهینه می‌شوند تا هم‌زمان دو هدف متضاد را برآورده سازند: حفظ ساختار محتوایی و بازتاب ویژگی‌های سبکی.

هدف اصلی الگوریتم انتقال سبک عصبی، یافتن تصویری  $x$  است که از یک سو محتوای تصویر  $p$  (مثلاً نمای یک شهر یا چهره‌ی انسان) را حفظ کند، و از سوی دیگر، سبک تصویر  $a$  (مثلاً نقاشی ون‌گوگ یا طرح موزاییکی) را بازتاب دهد. به این ترتیب، تصویر نهایی نوعی سازش میان دو فضا است: فضای ویژگی‌های ساختاری و فضای هم‌بستگی‌های آماری. تابع زیان کلی، به صورت ترکیب وزن‌دار دو بخش تعریف می‌شود:

$$(۱۲) \quad \mathcal{L}_{\text{total}}(p, a, x) = \alpha \mathcal{L}_{\text{content}}(p, x, l_c) + \beta \sum_{l \in \mathcal{L}} w_l \mathcal{L}_{\text{style}}(a, x, l)$$

که در آن:

- $l_c$  لایه‌ی محتوا و  $\mathcal{L}$  مجموعه‌ی لایه‌های سبک‌اند؛
- ضرایب  $\alpha$  و  $\beta$  میزان تأکید بر محتوا و سبک را تعیین می‌کنند.

از دیدگاه شهودی، این تابع نوعی مسئله‌ی بهینه‌سازی چندهدفه است که تعادلی میان «ساختار هندسی تصویر» و «آمار هم‌بستگی ویژگی‌ها» را با دو بخش «زیان محتوا» و «زیان سبک» برقرار می‌کند. زیان محتوا از اختلاف مستقیم ویژگی‌های استخراج‌شده از لایه‌ی خاصی از شبکه (مثلاً  $\text{conv4}_2$  در VGG-19) محاسبه می‌شود:

$$(۱۳) \quad \mathcal{L}_{\text{content}}(p, x, l) = \frac{1}{4} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2$$

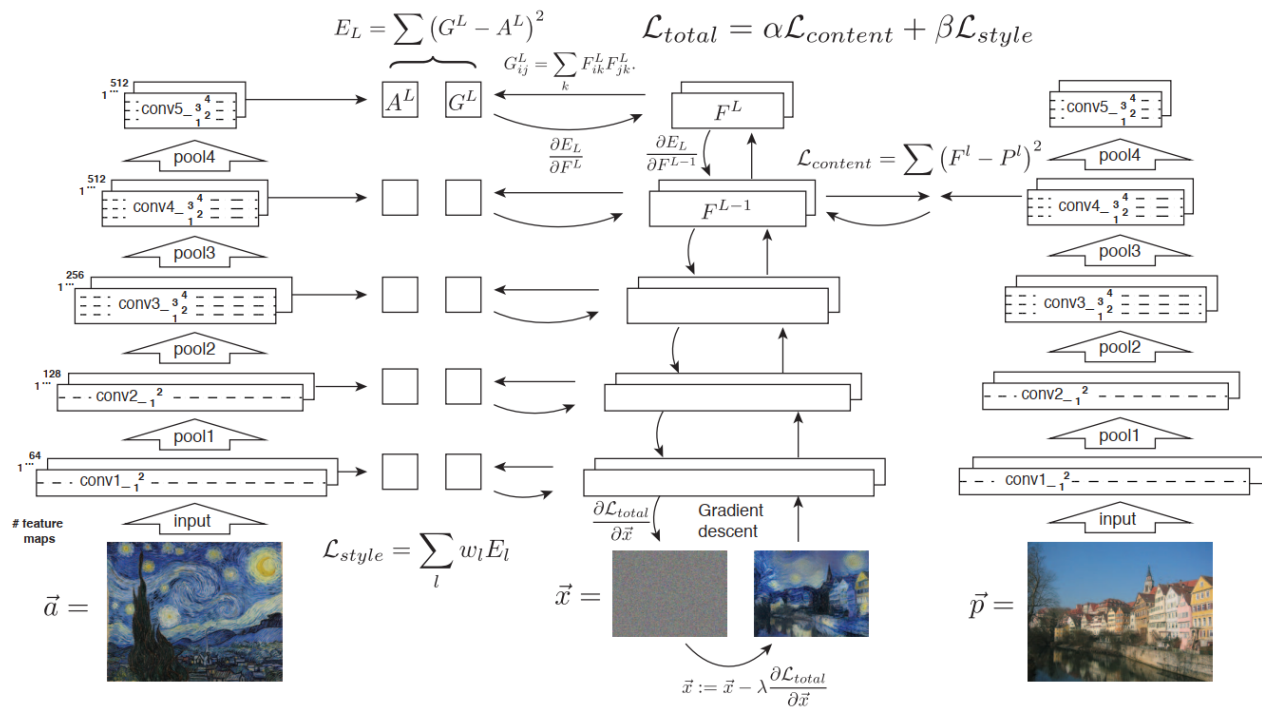
در این جا  $F_{ij}^l$  ویژگی‌های تصویر فعلی و  $P_{ij}^l$  ویژگی‌های تصویر محتوایی در لایه‌ی  $l$  هستند. زیان سبک بر پایه‌ی همان ماتریس گرامی تعریف می‌شود که در بخش قبل معرفی شد:

$$(۱۴) \quad \mathcal{L}_{\text{style}}(a, x, l) = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2$$

که در آن  $N_l$  تعداد فیلترها و  $M_l = H_l W_l$  تعداد نقاط مکانی در لایه‌ی  $l$  است. این معیار در واقع تفاوت میان ساختارهای هم‌بستگی ویژگی‌ها در تصویر فعلی و تصویر سبک را می‌سنجد. از منظر آماری، زیان محتوا سعی در حفظ میانگین ویژگی‌ها دارد، در حالی که زیان سبک به هم‌ترازی کواریانس‌ها می‌پردازد.

شکل ۱۱ چاقوب کلی الگوریتم انتقال سبک عصبی را نشان می‌دهد. در این الگوریتم، متغیر تصمیم، همان تصویر خروجی است، نه وزن‌های شبکه. شبکه‌ی CNN ثابت، نگه داشته می‌شود و تنها مقدار پیکسل‌های تصویر ورودی  $x$  به گونه‌ای تغییر می‌کند که تابع زیان کلی کمینه شود. فرآیند معمولاً با تصویری اولیه (تصویر محتوا یا نویز سفید) آغاز می‌شود و سپس گرادیان تابع زیان نسبت به پیکسل‌ها محاسبه می‌گردد:

$$(۱۵) \quad \frac{\partial \mathcal{L}_{\text{total}}}{\partial x} = \alpha \frac{\partial \mathcal{L}_{\text{content}}}{\partial x} + \beta \sum_l w_l \frac{\partial \mathcal{L}_{\text{style}}^l}{\partial x}$$



شکل ۱۱. شماتیک الگوریتم انتقال سبک [۹]: ویژگی‌های محتوایی و سبکی از تصاویر جداگانه استخراج می‌شوند و سپس تصویری جدید از طریق بهینه‌سازی گرادیانی به‌گونه‌ای ساخته می‌شود که هر دو نوع ویژگی را هم‌زمان بازتولید کند.

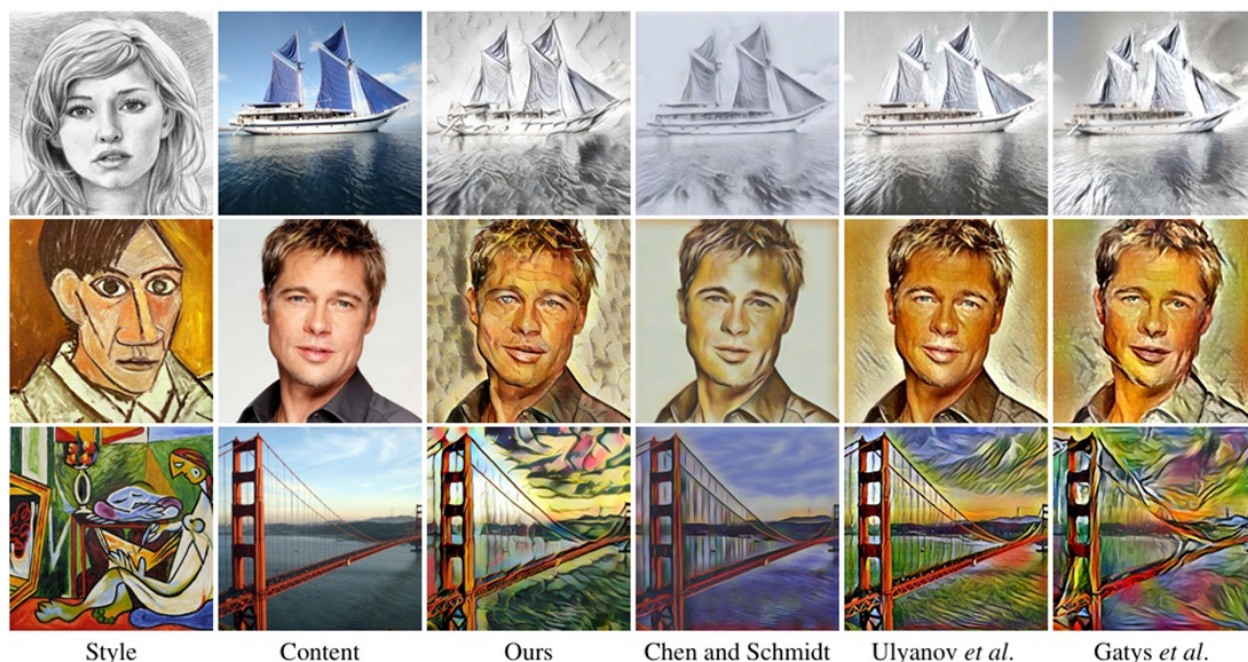
Figure 11: Schematic of the Neural Style Transfer algorithm [9]: content and style features are extracted from separate images, and then a new image is constructed through gradient optimization to simultaneously reproduce both types of features.

به‌روزرسانی پیکسل‌ها با روش‌های استاندارد بهینه‌سازی مانند <sup>۳۱</sup>L-BFGS، <sup>۳۲</sup>Adam یا <sup>۳۳</sup>SGD انجام می‌شود. نتیجه نهایی تصویری است که هم، ساختار تصویر محتوایی را بازتاب می‌دهد و هم بافت و رنگ‌آمیزی تصویر سبک (استایل) را در خود ترکیب کرده است. این الگوریتم نه‌تنها نقطه‌ی آغاز بسیاری از روش‌های مدرن انتقال سبک است، بلکه نمونه‌ای آموزنده از چگونگی ترکیب تفکر آماری و مفاهیم ریاضی با ساختارهای یادگیری عمیق به‌شمار می‌آید.

## ۶. از بهینه‌سازی تا نداشت‌های آماری: سیر تحول انتقال سبک

الگوریتم گاتیس و همکاران، هرچند از نظر مفهومی زیبا و از لحاظ ادراکی بسیار موفق بود، اما از دیدگاه محاسباتی پرهزینه است. محاسبه‌ی ماتریس‌های گرام برای هر لایه و اجرای صدها مرحله‌ی بهینه‌سازی برای هر تصویر، این روش را برای کاربردهای بلادرنگ نامناسب می‌کند. همین چالش‌ها انگیزه‌ای برای پژوهش‌های بعدی شدند تا با حفظ بنیان آماری روش اصلی، نسخه‌هایی سریع‌تر و عملی‌تر ارائه دهند [۱۵، ۱۶، ۲۴].

<sup>31</sup>Limited-memory Broyden-Fletcher-Goldfarb-Shanno <sup>32</sup>Adaptive Moment Estimation <sup>33</sup>Stochastic Gradient Descent



شکل ۱۲. بازنشر بخشی از [۱۲، شکل ۴]. مقایسه‌ی بصری، اصلی‌ترین روش مقایسه بین روش‌های مختلف انتقال سبک عصبی در مقالات هست.

Figure 12: A part of [12, Figure 4]. Visual comparison is the primary method of comparison among different Neural Style Transfer methods in the literature.

یکی از جهت‌گیری‌های کلیدی در این مسیر، جایگزینی بهینه‌سازی تکراری با نگاشت‌های مستقیم بود. در این رویکردها، به جای تغییر تدریجی پیکسل‌ها، یک شبکه‌ی پیش‌خور (فیدفورارد) آموزش داده می‌شود تا در یک عبور، تصویر خروجی را تولید کند. نمونه‌ی شاخص آن، کار جانسون و همکاران [۱۷] است که با استفاده از «زیان ادراکی» مبتنی بر ویژگی‌های CNN، زمان تولید سبک را از چند دقیقه به چند صدم ثانیه کاهش داد و کیفیت هنری را حفظ کرد. گام بعدی در این تحول، معرفی مفهوم «نرمال‌سازی تطبیقی»<sup>۳۴</sup> بود که توسط هوانگ و بلونگی ارائه شد [۱۲]. در این روش، به جای محاسبه‌ی ماتریس گرام، میانگین و انحراف معیار ویژگی‌های تصویر محتوا با مقادیر متناظر در تصویر سبک جایگزین می‌شوند. این ایده‌ی ساده اما مؤثر، انطباق آماری محتوا و سبک را بدون نیاز به بهینه‌سازی پیچیده ممکن می‌سازد و امکان انتقال سبک دلخواه را در زمان واقعی فراهم می‌کند. ویژگی مهم آن، این است که با یک شبکه‌ی واحد، کنترل‌هایی همچون تعادل میان محتوا و سبک، درونیابی میان چند سبک و تنظیم رنگ و موقعیت را فراهم می‌کند.

هم‌زمان، روش‌های دیگری مانند تبدیل سفید-رنگ‌آمیزی<sup>۳۵</sup> نیز توسعه یافتند [۲۱، ۱۵]. این روش با الهام از تبدیلات خطی مانند PCA، کوواریانس ویژگی‌های محتوا را سفید کرده و سپس با کوواریانس سبک «رنگ‌آمیزی» می‌کند تا سازگاری آماری مشابهی حاصل شود. در ادامه، کارهایی نظیر [۲۱] با افزودن زیان لاپلاسی، به حفظ جزئیات محتوایی و کاهش آثار تصنعی کمک کردند. بررسی‌های جامع [۱۵، ۱۶، ۲۴] نشان داده‌اند که انتقال سبک عصبی از یک پدیده‌ی هنری به

<sup>34</sup>adaptive instance normalization (AdaIN) <sup>35</sup>whitening and coloring transform



ابزاری عملی در حوزه‌های گوناگون تبدیل شده است؛ از طراحی و تولید محتوای هنری تا کاربردهای ویدئویی، مد، و واقعیت مجازی. با این وجود، چالش‌هایی مانند حفظ پیوستگی زمانی در ویدیوها و ارزیابی کمی کیفیت ادراکی همچنان از موضوعات باز و فعال پژوهشی به‌شمار می‌روند. در حال حاضر اصلی‌ترین روش مقایسه نتایج، مقایسه بصری روش‌های مختلف هست. به‌عنوان نمونه، شکل ۱۲ بخشی از [۱۲، شکل ۴] را نشان می‌دهد که نتایج چندین روش مورد مقایسه قرار گرفته‌اند.

موج دیگری از پژوهش‌ها در این حوزه، بر بهره‌گیری از مدل‌های پخش<sup>۳۶</sup> متمرکز شده است. این مدل‌ها با تولید تصویر از طریق حذف تدریجی نویز، توانسته‌اند کیفیت و تنوع بی‌سابقه‌ای در انتقال سبک ایجاد کنند. برای نمونه، روش «تزریق سبک در پخش»<sup>۳۷</sup> [۸] امکان انتقال سبک را بدون نیاز به بهینه‌سازی اضافی و تنها با دستکاری ویژگی‌های لایه‌های خودتوجهی در مدل‌های پخش بزرگ‌مقیاس فراهم می‌کند. رویکردهای مبتنی بر شبکه‌های مولد هم‌آورد (GAN)<sup>۳۸</sup> نیز همچنان با ارائه معماری‌های بهبودیافته، به‌ویژه برای حوزه‌های تخصصی مانند مصورسازی نقاشی‌ها [۲۶]، کارآمد و رقابتی باقی مانده‌اند. در تازه‌ترین تحولات، روش‌های مبتنی بر شبکه‌های برگشت‌پذیر<sup>۳۹</sup> مانند [۱۹] گامی فراتر از روش‌های پیشین برداشته و با استفاده از استراتژی تلفیق پارامترها، توانسته‌اند حفظ محتوا و وفاداری به سبک را همزمان بهبود بخشند. این تحولات نشان می‌دهد که روند تکامل انتقال سبک از روش‌های کلاسیک به سمت مدل‌های مولد عمیق، همچنان با شتاب در حال گسترش است.

به این ترتیب، مسیر تحول انتقال سبک عصبی (NST) از حل مستقیم یک مسئله بهینه‌سازی به سمت نگاشت‌های آماری و مدل‌های یادگیری‌پذیر حرکت کرده است. در تمامی این نسخه‌ها، جوهره ریاضی ثابت مانده است: هم‌ترازی آماری میان بازنمایی‌های لایه‌ای شبکه، که پلی میان پردازش سیگنال کلاسیک و یادگیری عمیق می‌سازد. در واقع، تابع هدف گاتیس را می‌توان نمونه‌ای از تلفیق دو حوزه بنیادین دانست: بهینه‌سازی تحلیلی در فضای ریاضی و یادگیری آماری در فضای ویژگی‌ها. به دنبال آن، روش‌های مبتنی بر زیان ادراکی [۱۷] و نرمال‌سازی آماری [۱۲] نشان دادند که حتی با ساده‌سازی‌های چشمگیر در ساختار ریاضی، می‌توان همان مفهوم توازن میان محتوا و سبک را در قالبی سریع‌تر و تعمیم‌پذیرتر حفظ کرد.

## ۷. جمع‌بندی و چشم‌انداز آینده

انتقال سبک عصبی را می‌توان نه صرفاً یک دستاورد در یادگیری عمیق، بلکه بازتفسیر مفاهیم بنیادین ریاضی در قالبی داده‌محور دانست. در این چارچوب، مفاهیمی چون پیچش، گرادیان و ماتریس کوواریانس – که ریشه در ریاضیات قرن بیستم دارند – در ساختار شبکه‌های عمیق بازآفرینی شده‌اند تا روابط آماری و هندسی نهفته در فضای تصویر را مدل‌سازی کنند. در سال‌های اخیر، ایده بنیادی انتقال سبک عصبی الهام‌بخش شاخه‌های تازه‌ای در هنر محاسباتی شده است. در حوزه «نقاشی عصبی»<sup>۴۰</sup>، همین چارچوب نظری برای شبیه‌سازی فرایند نقاشی با قلم‌مو و ترکیب رنگ‌ها به‌کار گرفته می‌شود [۱۳، ۳۰]. همچنین روش‌هایی چون CartoonGAN و CycleGAN با گسترش مفهوم هم‌ترازی آماری و ویژگی‌ها، امکان بازآفرینی هدفمند سبک‌های خاصی مانند نقاشی، کمیک یا انیمه را فراهم کرده‌اند [۷] و در مسائل شناسایی به‌عنوان راهکاری برای داده‌افزایی مورد استفاده قرار گرفته‌اند [۵]. این روند نشان می‌دهد که ایده گاتیس و همکاران افق‌های نوینی در مرز میان هنر و یادگیری عمیق گشوده است.

امروزه انتقال سبک از مرزهای هنر فراتر رفته و در حوزه‌هایی چون پزشکی تصویری، طراحی صنعتی، واقعیت مجازی، فشرده‌سازی داده و تحلیل زیبایی‌شناسی به‌کار گرفته می‌شود. با وجود این پیشرفت‌ها، چالش‌های نظری و محاسباتی متعددی همچنان باقی هستند؛ از جمله تعریف دقیق‌تر فضای سبک، طراحی متریک‌های ادراکی هم‌راستا با قضاوت انسانی و توسعه مدل‌هایی که بتوانند چندین سبک را به‌صورت سلسله‌مراتبی و تفسیرپذیر ترکیب کنند. افزون بر این، پژوهش‌های نوین در حال

<sup>36</sup>Diffusion Models    <sup>37</sup>Style Injection in Diffusion    <sup>38</sup>Generative Adversarial Networks (GAN)    <sup>39</sup>reversible networks

<sup>40</sup>neural painting

بررسی پیوند میان شبکه‌های عصبی و ابزارهای تحلیل چندمقیاسی مانند تبدیلات موجک و لاپلاسی هستند، تا تفسیرپذیری ریاضی را با توان یادگیری داده‌محور در یک چارچوب واحد ادغام نمایند.

در نگاهی کلان، مسیر تحول انتقال سبک عصبی بازتابی از حرکت گسترده‌تر در علم پردازش تصویر است: گذار از مدل‌های تحلیلی و نظری کلاسیک به سوی سامانه‌های یادگیرنده و داده‌محور؛ حرکتی که در عین حال، پیوند دیرینه‌ی میان ریاضی، آمار، علوم کامپیوتر و هنر را زنده نگاه می‌دارد.

## مراجع

- [1] M. Amintoosi, M. Fathy and N. Mozayani, Regional varying image super-resolution, *In: IEEE international joint conference on computational sciences and optimization*, **1** (2009), 913–917.
- [2] M. Amintoosi and F. Farbiz, Eigenbackground revisited: can we model the background with eigenvectors?, *J. Math. Imaging Vision*, **64** (2022), no. 5, 463–477.
- [3] M. Amintoosi, Application of Taylor expansion in reducing the size of convolutional neural networks for classifying Impressionism and Miniature paintings, *Math. Soc.*, **5** (2020), no. 1, 1–16. [In Persian]. [https://math-sci.ui.ac.ir/article\\_25351.html](https://math-sci.ui.ac.ir/article_25351.html).
- [4] M. Amintoosi, Fully connected to fully convolutional: road to yesterday, *Journal of soft computing and information technology*, **11** (2022), no. 1, 60–72. [In Persian]. [https://jscit.nit.ac.ir/article\\_149453.html](https://jscit.nit.ac.ir/article_149453.html).
- [5] M. Amintoosi, Style transfer for data augmentation in convolutional neural networks applied to fire detection, *Computational intelligence in electrical engineering*, **13** (2022), no. 4, 97–114. [In Persian]. [https://isee.ui.ac.ir/article\\_26042.html](https://isee.ui.ac.ir/article_26042.html).
- [6] T. Y. Chiu and D. Gurari, PCA-based knowledge distillation towards lightweight and content-style balanced photorealistic style transfer models, *In: 2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Los Alamitos, CA, USA: IEEE Computer Society, (2022), 7834–7843, <https://doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.00769>.
- [7] Y. Chen, Y. K. Lai and Y. J. Liu, CartoonGAN: generative adversarial networks for photo cartoonization, *In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, (2018), 9465–9474. <https://doi.org/10.1109/CVPR.2018.00986>.
- [8] J. Chung, S. Hyun and J. P. Heo, Style injection in diffusion: a training-free approach for adapting large-scale diffusion models for style transfer, *In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, (2024), 8795–8805.
- [9] L. A. Gatys, A. S. Ecker and M. Bethge, Image style transfer using convolutional neural networks, *In: 2016 IEEE conference on computer vision and pattern recognition (CVPR)*, (2016), 2414–2423.
- [10] L. A. Gatys, A. S. Ecker and M. Bethge, Texture synthesis using convolutional neural networks, *In: Proceedings of the 29th international conference on neural information processing systems*, **1**, NIPS'15, Cambridge, MA, USA: MIT Press; (2015), 262–270.
- [11] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, (2016).

- [12] X. Huang and S. Belongie, Arbitrary style Transfer in real-time with adaptive instance normalization, *In: Proceedings of the IEEE international conference on computer vision (ICCV)*, (2017), 1510–1519. <https://doi.org/10.1109/ICCV.2017.167>.
- [13] A. Hertzmann, Can computers create art?, *Arts*, **7** (2018), no. 2, pp. 25. <https://www.mdpi.com/2076-0752/7/2/18>.
- [14] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu and M. Song, Neural style transfer: a review, *IEEE Transactions on Visualization & Computer Graphics*, **26** (2020) , no. 11, 3365–3385. <https://doi.org/10.48550/arXiv.1705.04058>.
- [15] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu and M. Song, Neural style transfer: a review, *IEEE transactions on visualization and computer graphics*, **26** (2020), no. 11, 3365–3385. <https://doi.org/10.1109/TVCG.2019.2921336>.
- [16] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu and M. Song, *Neural style transfer: a review*, (updated 2022 version), (2022). arXiv preprint arXiv:1705.04058. <https://arxiv.org/abs/1705.04058>.
- [17] J. Johnson, A. Alahi and L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, *In: Proceedings of the european conference on computer vision (ECCV)*, (2016), 694–711. [https://doi.org/10.1007/978-3-319-46475-6\\_43](https://doi.org/10.1007/978-3-319-46475-6_43)
- [18] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *In: F. Pereira, Burges CJC, L. Bottou, K. Q. Weinberger, editors, Advances in neural information processing systems 25*, Curran Associates, Inc., (2012) 1097–1105.
- [19] Y. T. Kao, H. J. Lin, K. J. Lin and Y. Tokuyama, Content-preserving image style transfer via reversible networks with meta ActNorm, *Electronics*, **15** (2026), no. 2, 395. <https://doi.org/10.3390/electronics15020395>.
- [20] X. Li, S. Liu, J. Kautz and M. H. Yang, Learning linear transformations for fast image and video style transfer, *In: IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, (2019), 3804–3812.
- [21] S. Li, X. Xu, L. Nie and T. S. Chua, Laplacian-steered neural style transfer, *In: Proceedings of the 26th international conference on world wide web (WWW)*, (2017), 1123–1132. <https://doi.org/10.1145/3038912.3052617>.
- [22] J. Redmon and A. Farhadi, YOLO9000: Better, Faster, Stronger, CVPR, 2017.
- [23] K. Simonyan and A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, *In: International conference on learning representations*, (2015).
- [24] A. Singh, V. Jaiswal, G. Joshi, A. Sanjeev, S. Gite and K. Kotecha, Neural style transfer: a critical review, *IEEE Access*, **9** (2021) ,60297–60311. <https://doi.org/10.1109/ACCESS.2021.3112996>.
- [25] A. Vidhya, The world through the eyes of CNN, 2020. Accessed: (2025). <https://medium.com/analytics-vidhya/the-world-through-the-eyes-of-cnn-5a52c034dbeb>.
- [26] X. Wang and J. Jiang, Illustration image style transfer method design based on improved cyclic consistent adversarial network, *PLOS ONE*, **20** (2025), no. 1, e0313113.

- [27] J. Yoo, Y. Uh, S. Chun, B.Kang and J. W. Ha, Photorealistic style transfer via wavelet transforms, *In: IEEE/CVF international conference on computer vision (ICCV)*, (2019), 9035–9044.
- [28] M. D. Zeiler and R. Fergus, Visualizing and understanding convolutional networks, *In: European conference on computer vision (ECCV)*, Springer, (2014), 818–833.
- [29] Y. Zheng, C. Yang and A. Merkulov, Breast cancer screening using convolutional neural network and follow-up digital mammography, In: A. Mahalanobis, A. Ashok, L. Tian and J. C. Petrucci, editors, *Conference: Computational Imaging III*, **10669**, International society for optics and photonics, SPIE, (2018). <https://doi.org/10.1117/12.2304564>.
- [30] Z. Zou, T. Shi, S. Qiu, Y. Yuan and Z. Shi, Stylized neural painting, *In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Los Alamitos, CA, USA: IEEE Computer Society, (2021), 15684–15693. <https://doi.ieeecomputersociety.org/10.1109/CVPR46437.2021.01543>.

محمود امین طوسی

گروه ریاضی کاربردی، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد، مشهد، ایران.

m.amintoosi@um.ac.ir

محمود امین طوسی، دانشیار علوم کامپیوتر دانشگاه فردوسی مشهد است. وی دوره‌های کارشناسی و کارشناسی ارشد خود را در رشته‌های ریاضی (گرایش کاربرد در کامپیوتر) و مهندسی کامپیوتر (گرایش نرم‌افزار) در دانشگاه فردوسی به اتمام رسانده و دوره دکترای خود را در رشته مهندسی کامپیوتر (گرایش هوش مصنوعی) در دانشگاه علم و صنعت ایران گذرانده است. علائق پژوهشی وی یادگیری ماشین، بینایی ماشین و بهینه‌سازی ترکیبیاتی می‌باشد. اسلایدهای مرتبط با این مقاله از بخش Presentations از صفحه شخصی نگارنده به آدرس <https://mamintoosi.github.io/> قابل دسترس می‌باشد.

