

ارائه یک روش یادگیری خود - نظارتی عمیق مبتنی بر تبدیل موجک گسسته دو بعدی برای تعمیم دامنه تصاویر

سارا فرهمندی نیا^۱، مهدی افتخاری^۲ و کاوه بهرامن^۳

چکیده

در یادگیری ماشین، انتقال و تعمیم دانش یادگرفته شده از یک دامنه به دامنه‌های دیگر، یکی از قابلیت‌های مهم و اساسی به شمار می‌رود. از آنجا که یادگیری با نظارت هرگز نمی‌تواند کامل باشد، استفاده از روش‌های دیگری همچون روش‌های یادگیری خود - نظارتی می‌تواند برای مسأله‌ی تعمیم دامنه بسیار کمک‌کننده باشد. در این مقاله، ما روشی را ارائه می‌دهیم که علاوه بر طبقه بندی تصاویر اصلی به منظور یادگیری برچسب‌های داده در فرایند با نظارت، سعی می‌کند که تصاویر حاصل از اعمال تبدیل موجک گسسته بر روی تصاویر اصلی را با تولید شبه برچسب‌هایی برای آنها طبقه بندی کند. این کار به عنوان یک وظیفه‌ی خود - نظارتی می‌تواند باعث یادگیری ویژگی‌های مفید و یک بازنمایش کلی در میان تصاویر دامنه‌های مختلف شود، که می‌تواند به بهبود مسأله‌ی تعمیم دامنه بسیار کمک کند. در ادامه با ترکیب روش‌های خود - نظارتی مانند پازل jigsaw و حدس زاویه چرخش با تبدیل موجک گسسته، نشان می‌دهیم که این ترکیب می‌تواند باعث بهبود نتایج برای مسأله‌ی تعمیم دامنه شود. در این مقاله، ما از مجموعه داده‌های معروف PACS، VLCS و Office-Home برای انجام آزمایش‌ها استفاده کردیم و نتایج نشان می‌دهند که روش پیشنهادی ما می‌تواند از روش‌های پیشرفته و به روز تعمیم دامنه بهتر عمل کند.

کلیدواژه‌ها

تطبیق دامنه، تعمیم دامنه، دامنه منبع، دامنه هدف، یادگیری - خودنظارتی، تبدیل موجک.

۱ مقدمه

در یادگیری ماشین، مسائل با توجه به داده‌هایی که در اختیار قرار این مقاله در دی‌ماه ۹۹ دریافت، در تیرماه ۱۴۰۰ بازنگری دوم و در مردادماه همان سال پذیرفته شد.

۱ بخش مهندسی کامپیوتر، دانشگاه شهید باهنر کرمان.

رایانامه: sara.farahmandi90@gmail.com

۲ بخش مهندسی کامپیوتر، دانشگاه شهید باهنر کرمان.

رایانامه: m.eftekhari@uk.ac.ir

۳ بخش مهندسی کامپیوتر، دانشگاه شهید باهنر کرمان.

رایانامه: sadegh.b218@gmail.com

نویسنده مسئول: مهدی افتخاری

DOR: 20.1001.1.23831197.1401.9.1.5.6

دارند به دو دسته‌ی اصلی یادگیری با نظارت^۱ و یادگیری بدون نظارت^۲ تقسیم بندی می‌شوند. در یادگیری با نظارت مجموعه‌ای از جفت‌های ورودی - خروجی (داده‌های با برچسب^۳) به سیستم داده می‌شود و سیستم تلاش می‌کند تا تابعی از ورودی به خروجی را فراگیرد. اما در یادگیری بدون نظارت، یادگیری بر روی داده‌های بدون برچسب و برای یافتن الگوهای پنهان در این داده‌ها انجام می‌گیرد [۲۵].

یادگیری با نظارت در کارهایی مانند طبقه بندی تصاویر^۴ و تشخیص اشیاء^۵ پیشرفت چشمگیری داشته است. در مسأله‌ی

¹ Supervised learning

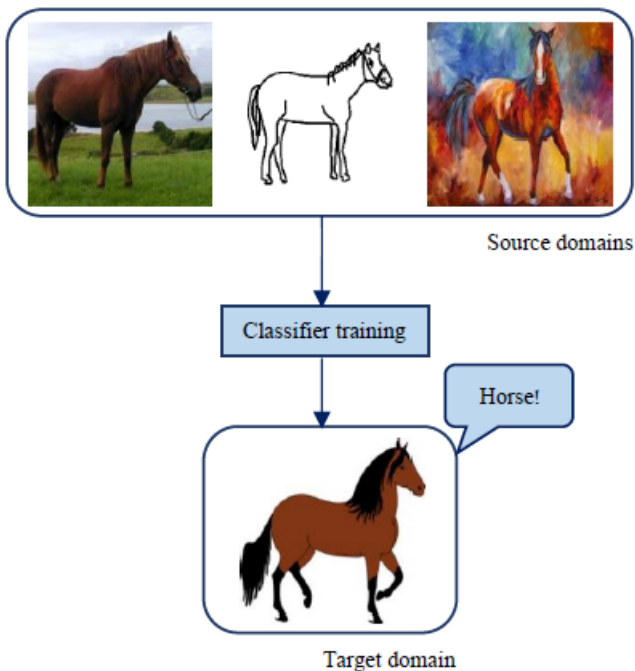
² Unsupervised learning

³ lable

⁴ Image classification

⁵ Object detection

دیگری از همان شیء در دامنه‌ی هدف جدیدی که در زمان آموزش از آن برخوردار نبوده است را به خوبی تشخیص دهد. تعمیم دامنه و تطبیق دامنه دو روش مجزا ولی نزدیک به هم هستند که هر دو وظیفه اختصاص برچسب به یک مجموعه بدون برچسب را بر عهده دارند. با این تفاوت که در تطبیق دامنه، داده‌های هدف در مرحله آموزش در دسترس قرار دارند اما در تعمیم دامنه، داده‌های هدف در زمان آموزش دیده نمی‌شوند. روش‌های تطبیق دامنه‌ی معمولی بدون داشتن داده‌های هدف در مرحله‌ی آموزش عملکرد خوبی ندارند، زیرا اگر در روش‌های تطبیق دامنه به طور مستقیم داده‌های هدف را از آموزش حذف کنیم، عملکرد ضعیفی حاصل می‌شود. تعمیم دامنه موضوعی است که کمتر از تطبیق دامنه مورد مطالعه قرار گرفته و چالش برانگیز است زیرا دانش قبلی در مورد دامنه هدف وجود ندارد [۱۲]. در ادامه به منظور روشن تر شدن مسأله، تعریفی از دامنه و تعمیم دامنه با فرمول بندی ارائه می‌شود.



شکل ۱ تعمیم دامنه میان دامنه‌های منبع گوناگون.

تعریف ۱ (دامنه): فرض کنید X یک فضای ورودی غیرتهی و Y یک فضای خروجی باشد. یک دامنه از داده‌هایی تشکیل شده است که از یک توزیع نمونه برداری می‌شوند. دامنه به صورت $s = \{(x_i, y_i)\}_{i=1}^n \sim \rho_{xy}$ تعریف می‌شود به طوری که: $x \in X \subset R^d, y \in Y \subset R$ و ρ_{xy} توزیع مشترک^۷ نمونه ورودی x و برچسب خروجی y را نشان می‌دهد [۲۷].

تعریف ۲ (تعمیم دامنه): در مسأله‌ی تعمیم دامنه، M دامنه (منبع) برای آموزش داده می‌شود:

$$S_{train} = \{s^i \mid i = 1, \dots, M\}$$

طبقه بندی، بر اساس داده‌های قبلی که دارای برچسب هستند مدلی برای پیش بینی برچسب داده‌های جدید ساخته می‌شود و به صورت استاندارد، یک مدل بر روی داده‌های آموزش یا منبع، آموزش می‌بیند و سپس بر روی داده‌های تست یا هدف، برای پیش بینی اعمال می‌شود. ممکن است به صورت ضمنی تصور شود که داده‌های منبع و داده‌های هدف دارای توزیع یکسان باشند، در حالی که در دنیای واقعی این فرض رد می‌شود و ممکن است داده‌های منبع و داده‌های هدف از منابع مختلف و تحت شرایط مختلف (به عنوان مثال: نماها^۱، نور، پس زمینه و غیره که هر کدام را می‌توان به عنوان یک دامنه در نظر گرفت) جمع آوری شوند [۲۶].

در برخی از کاربردهای بینایی رایانه، مواردی وجود دارد که برای یک دامنه‌ی مورد نظر که به عنوان دامنه هدف^۲ نام برده می‌شود، داده‌های موجود، محدود و بدون برچسب هستند و بنابراین داده آموزش کافی وجود ندارد، در حالی که داده‌های دارای برچسب فراوانی در دامنه‌های مرتبط با دامنه هدف (به عنوان دامنه‌های منبع^۳) وجود دارد که می‌تواند بسیار کمک کننده باشد. در چنین مواردی، با توجه به این که جمع آوری و حاشیه نویسی مجموعه داده‌ها برای هر دامنه‌ی جدید کار دشواریست، از داده‌های دامنه‌های مرتبط برای آموزش استفاده می‌شود. اما به دلیل تفاوت دامنه، ممکن است مدلی که با داده‌های منبع آموزش دیده است نتواند عملکرد مناسبی برای دامنه هدف داشته باشد. برای رفع این مشکل، اخیراً مطالعات بسیاری در پیشبرد تطبیق مدل آموخته شده با استفاده از داده‌های دارای برچسب منبع، با دامنه هدف انجام شده است. به این روش‌ها، روش‌های تطبیق دامنه^۴ گفته می‌شود [۱۱] و [۱۴].

در روش‌های تطبیق دامنه در مرحله‌ی آموزش دسترسی به داده‌های دامنه‌ی هدف وجود دارد و این در حالیست که در مشکلات دنیای واقعی ممکن است داده‌های هدف در زمان آموزش در دسترس نباشند، اما هنوز یک مدل دقیق برای دامنه هدف نادیده نیاز باشد. برای حل این مشکل روش‌های تعمیم دامنه^۵ مطرح می‌شوند، که با استفاده از داده‌های دارای برچسب دامنه‌های منبع، نمایه^۶ عمومی را یاد می‌گیرند که انتظار می‌رود برای دامنه هدف به خوبی تعمیم یابد [۱۱]. شکل ۱، مسأله‌ی تعمیم دامنه میان دامنه‌های منبع گوناگون را نشان می‌دهد. همان طور که در شکل مشاهده می‌شود، تصاویر مختلفی از یک شیء در دامنه‌های گوناگون (تصاویری که تحت شرایط مختلف ایجاد شده‌اند و دارای پس زمینه، زاویه و بافت متفاوتی نسبت به یکدیگر هستند) به یک طبقه بند داده می‌شود و انتظار می‌رود مدل ایجاد شده بتواند تصویر

¹ Viewpoints

² Target domain

³ Source domain

⁴ Domain adaptation

⁵ Domain generalization

⁶ representation

⁷ Joint distribution

jigsaw، حدس زاویه چرخش تصویر، بازسازی تصویر آسیب دیده^{۱۵} و غیره می باشد [۱۶].

تبدیل موجک^{۱۶} یک تکنیک قدیمی و مشهور در پردازش سیگنال است که در پردازش تصویر نیز کاربرد دارد و ویژگی های مهم مانند لبه ها^{۱۷}، نواحی روشن و ناهنجاری های دیگر را استخراج می کند. تبدیل موجک در زمینه های متفاوتی مورد استفاده قرار می گیرد و کاربردهای گوناگونی دارد [۲۱]. این تبدیل دارای دو نوع مختلف پیوسته و گسسته است. هنگامی که تبدیل موجک گسسته بر روی یک تصویر که یک سیگنال دو بعدی است، اعمال می شود، تصویر را به چهار زیرباند مختلف تجزیه می کند که ویژگی های مربوط به کلیات تصویر، جزئیات افقی، عمودی و قطری تصویر را استخراج می کنند. شکل ۴ در قسمت ۳، زیرباندهای مختلف حاصل از تبدیل موجک یک تصویر را نشان می دهد. همان طور که در این شکل مشاهده می شود، ویژگی های استخراج شده بیشتر بر روی ساختار کلی شیء متمرکز هستند، به همین دلیل یادگیری چنین ویژگی هایی می تواند برای مسأله ی تعمیم دامنه مفید واقع شود.

در این مقاله، ما یک روش جدید برای تعمیم دامنه ارائه می دهیم که از تبدیل موجک گسسته دو بعدی به صورت یک روش یادگیری خود - نظارتی استفاده می کند و یاد می گیرد که چگونه در میان دامنه های مختلف تعمیم یابد. در این روش ما از یک شبکه عصبی یکسان به منظور یادگیری همزمان برچسب های داده از تصاویر اصلی در یک حالت با نظارت و یادگیری ویژگی های عمومی و مهم از تصاویر مختلف حاصل از تبدیل موجک در یک حالت خود - نظارت استفاده می کنیم. در این مقاله، ما نشان می دهیم که این روش می تواند از بسیاری از روش های موجود برای مسأله ی تعمیم دامنه بهتر عمل کند. در ادامه نشان می دهیم که استفاده از ترکیبی از روش های یادگیری خود - نظارتی همانند پازل jigsaw، حدس زاویه چرخش و تبدیل موجک می تواند نتایج بسیار خوبی را در برداشته باشد.

۲ کارهای پیشین

۲-۱ تعمیم دامنه

در مسأله ی تعمیم دامنه، چالش اصلی یادگیری اطلاعات مفید عمومی و قابل انتقال در میان دامنه های مختلف است و تاکنون روش های متعددی برای آن ارائه شده است.

روش های مبتنی بر یادگیری نمایه از محبوب ترین روش ها برای تعمیم دامنه محسوب می شوند. در این روش ها طبقه بندی با استفاده از ویژگی های استخراج شده به وسیله یک تابع، انجام می شود و بسیاری روش ها برای یادگیری بهتر این تابع، که تابع

به طوری که $s^i = \{(x_j^i, y_j^i)\}_{j=1}^{n_i}$ ، i امین دامنه را نشان می دهد و توزیع های مشترک میان هر دو دامنه متفاوت است یعنی:

$$\rho_{xy}^i \neq \rho_{xy}^j, 1 \leq i \neq j \leq M$$

هدف از تعمیم دامنه، یادگیری یک مدل پیش بینی قوی و قابل تعمیم $h: X \rightarrow Y$ از M دامنه ی آموزش، برای دستیابی به حداقل خطای پیش بینی اعمال شده برای دامنه ی هدف نادیده Stest است که در زمان آموزش در دسترس نیست یعنی:

$$\min_h E_{(x,y) \in S_{test}} [l(h(x), y)]$$

به طوری که $\rho_{xy}^{test} \neq \rho_{xy}^i$ for $i \in \{1, \dots, M\}$ و تابع انتظار^۱ و L تابع هزینه^۲ هستند [۲۷].

از کارهای قبلی در مورد تعمیم دامنه که از گذشته تاکنون مطرح شده اند می توان از روش های کم عمق^۳ [۱] تا [۸]، روش های مبتنی بر توسعه رویکردهای داده محور برای یادگیری ویژگی های ثابت^۴ در دامنه های منبع مختلف [۲] و [۴] و [۳] و [۹] و [۱۰]، روش های مبتنی بر معماری عمیق^۵ از جمله استفاده از آموزش خصمانه^۶ برای یادگیری ویژگی های نهفته^۷ عمومی در میان میان دامنه ها [۱۱] و [۱۵]، تولید کردن نمونه های جدید به منظور کم کردن فاصله بین دامنه های منبع با دامنه هدف [۱۲] و استفاده از یک روش یادگیری خود - نظارتی^۸ با هدف یادگیری ویژگی های ثابت در میان دامنه های مختلف [۱۳] نام برد.

تاکنون روش های یادگیری با نظارت بسیاری برای مسأله ی تعمیم دامنه ارائه شده است، اما به دلیل این که داده های بدون برچسب در حجم بالایی در دسترس هستند و موضوع گرایش^۹ به برچسب داده ها مطرح نمی شود، به نظر می رسد که یادگیری بدون نظارت گزینه خوبی برای یادگیری اطلاعات مستقل از دامنه باشد [۱۳]. یادگیری خود - نظارتی که زیرمجموعه ای از یادگیری بدون نظارت است، با ایجاد یک کار بهانه^{۱۰} (به عنوان مثال حدس زاویه چرخش^{۱۱} تصویر) و تولید شبه برچسب^{۱۲} های مرتبط با آن کار برای مجموعه ای از داده های بدون برچسب، سعی می کند مدلی را با هدف یادگیری ویژگی های عمومی نهفته آموزش دهد، که می تواند برای یک کار پایین دستی^{۱۳} (مانند طبقه بندی تصاویر یا تشخیص اشیاء) مفید واقع شود. یادگیری خود - نظارتی دارای روش های متعددی از جمله رنگ آمیزی تصویر^{۱۴}، حل پازل

¹ Expectation function

² Loss function

³ shallow

⁴ Invariant features

⁵ Deep structure

⁶ Adversarial

⁷ Latent features

⁸ Self - supervised

⁹ bias

¹⁰ Pretext task

¹¹ rotation

¹² Pseudo - label

¹³ Down - stream task

¹⁴ Image colorization

¹⁵ inpainting

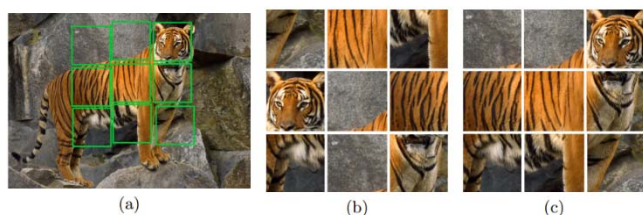
¹⁶ Wavelet transform

¹⁷ edges

سیگنال‌های خود - نظارت با چگونگی حل پازل jigsaw در همان تصاویر درک خود از داده‌ها را گسترش می‌دهد. این کار به شبکه کمک می‌کند تا مفاهیم همبستگی مکانی را در حالی که به عنوان تنظیم کننده برای کار طبقه بندی عمل می‌کند، بیاموزد. روش پیشنهادی ما مبتنی بر این قالب است و در ادامه، این روش به عنوان یک معیار برای ارزیابی انتخاب شده است.

۲-۲ یادگیری خود - نظارتی

همان طور که گفته شد، در یادگیری خود - نظارتی هدف آموزش یک مدل با استفاده از شبه برچسب‌های تولید شده از یک کار بهانه، به منظور یادگیری نمایه‌های مفید برای یک کار پایین دستی است. حل پازل jigsaw یک روش یادگیری خود - نظارتی است که به عنوان یک کار بهانه به منظور یادگیری نمایه بصری بدون نظارت کاربردهای بسیاری دارد. حل پازل jigsaw به معنای قطعه قطعه کردن یک تصویر و بهم ریختن قطعه‌ها^{۱۱} و سپس ایجاد تصویر اصلی با استفاده از قطعه‌های به هم ریخته، می‌باشد. شکل ۲، یادگیری نمایه‌های تصویر به وسیله حل پازل jigsaw را نشان می‌دهد. در [۱۷] روشی برای تشخیص جابه جایی^{۱۲} قطعه‌های پازل jigsaw ارائه شده است.



شکل ۲ یادگیری نمایه‌های تصویر به وسیله حل پازل jigsaw [۱۷].

به منظور بهبود ویژگی‌های آموخته شده برای کار پایین دستی روش‌های دیگری مبتنی بر حل پازل jigsaw [۱۸] نیز مطرح شده‌اند. در روش JiGen [۱۳] از پازل jigsaw به صورت یک مسأله طبقه بندی استفاده می‌کند، به این صورت که به هر ترتیب از قطعه‌های پازل یک برچسب اختصاص می‌دهد، سپس شبکه را مجبور می‌کند تا برچسب تصویر بهم ریخته^{۱۳} را پیش بینی کند. حدس زاویه چرخش تصویر نیز یکی دیگر از روش‌های یادگیری خود - نظارتی است، که تصویر را با یک زاویه تصادفی چرخانده و از شبکه خواسته می‌شود تا زاویه چرخش تصویر را پیش بینی کند. این روش در [۱۹] به عنوان روشی برای یادگیری نمایه بصری بدون نظارت ارائه شده است. از ترکیب روش‌های فوق در [۲۰] به عنوان روشی جدید برای تطبیق دامنه استفاده شده که نشان می‌دهد ترکیب این روش‌ها می‌تواند در مواردی باعث یادگیری بهتر نمایه ویژگی‌های تصاویر شود.

استخراج ویژگی یا تابع یادگیری نمایه نامیده می‌شود، طراحی می‌شوند. روش‌های مبتنی بر هسته^۱ [۱] و [۲] روش‌هایی هستند که در آن‌ها تابع یادگیری نمایه به عنوان یک نگاشت ویژگی^۲ پیاده سازی می‌شود که به آسانی با استفاده از یک تابع هسته (مانند تابع پایه شعاعی و لاپلاس) محاسبه می‌شود. M. Ghifary و همکاران روشی را با استفاده از خودرزمگذارهای^۳ چندمنظوره^۴ [۴] و S. Motiian و همکاران روشی را با استفاده از یک تابع هزینه مقابله-ای^۵ [۹] ارائه کردند که، از طریق به حداقل رساندن اختلاف میان دامنه‌های منبع که پایه و اساس روش‌های مبتنی بر هم تراز^۶ دامنه دامنه است، سعی می‌کنند تا نمایه ثابت میان آنها را یاد بگیرند. آموزش خصمانه نیز، به طور گسترده‌ای برای یادگیری ویژگی‌های ثابت در میان دامنه‌ها مورد استفاده قرار می‌گیرد و روش‌هایی ارائه شده است که از یک شبکه خصمانه به منظور یادگیری ویژگی‌های نهفته عمومی در میان دامنه‌های مختلف، برای تعمیم دامنه استفاده می‌کنند [۱۱] و [۱۵]. روش‌های مبتنی بر یادگیری نمایه دیگری نیز وجود دارند که تجزیه و تحلیل چند مؤلفه‌ای^۷ نامیده می‌شوند و سعی می‌کنند تا با استفاده از پارامترهای شبکه، ویژگی‌های مشترک دامنه‌ها و ویژگی‌های خاص هر دامنه را یاد بگیرند [۳] و [۵] و [۷] و [۱۰].

گروهی دیگر از روش‌هایی که برای تعمیم دامنه مطرح شده است، روش‌های مبتنی بر دستکاری^۸ داده‌ها هستند که سعی می‌کنند تا با افزایش تنوع داده‌های موجود برای آموزش، توانایی تعمیم مدل را تقویت کنند. Mohammad Mahfujur Rahman و همکاران روشی را ارائه کردند که به وسیله یک شبکه GAN داده‌های مصنوعی تولید می‌کند و از این قبیل روش‌ها به حساب می‌آید [۱۲]. علاوه بر این روش‌ها، تعمیم دامنه همچنین در پارادایم‌های کلی یادگیری ماشین به عنوان مثال، روش‌های مبتنی بر یادگیری گروهی^۹ [۲۸]، روش‌های مبتنی بر فرا یادگیری^{۱۰} [۲۹] و [۳۰] و روش‌های مبتنی بر یادگیری خودنظارتی [۱۳] مورد مطالعه قرار گرفته است.

برخی از روش‌هایی که تاکنون ذکر شد، از ویژگی‌های کم عمق استفاده می‌کنند و برخی نیز مبتنی بر ساختارهای عمیق هستند. اخیراً در زمینه تعمیم دامنه Fabio M. Carlucci و همکاران [۱۳] روشی به نام JiGen ارائه کردند که از یک روش یادگیری خود - نظارتی به منظور یادگیری ویژگی‌های عمومی یکسان در میان دامنه‌ها استفاده می‌کند. این روش برچسب‌های تصاویر را به صورت با نظارت یاد می‌گیرد و همزمان، با یادگیری

1 kernel
2 Feature map
3 autoencoders
4 Multitask
5 contrastive
6 alignment
7 Multi-component analysis
8 manipulation
9 Ensemble learning
10 Meta learning

11 patches
12 permutation
13 Shuffled image

۲-۳ تبدیل موجک

تبدیل موجک یک ابزار ریاضی مشهور برای تجزیه و تحلیل زمان - فرکانس است، مجموعه‌ای از موجک‌ها از موجک اصلی مادر از طریق تأخیر و شیفت عملیات تولید می‌شوند که می‌توانند برای تحلیل ویژگی‌های متفاوت سیگنال به کار روند. در خصوص پردازش تصویر، نسبت فرکانس بالا، جزئیات کوچک مقیاس مانند لبه‌های تصاویر را نشان می‌دهد، در حالی که نسبت فرکانس پایین، مربوط به قسمت صاف و هموار تصاویر است [۲۲]. تبدیل موجک در زمینه‌هایی مانند ریاضیات، پردازش سیگنال، تشخیص الگو، پردازش تصویر و گرافیک رایانه‌ای و تصویربرداری پزشکی استفاده می‌شود. تبدیل موجک کاربردهای وسیعی دارد، از آن جمله می‌توان به فشرده سازی تصویر^۱ [۲۳]، کاهش نویز^۲ [۲۴]، قطعه بندی تصویر^۳ [۲۲]، طبقه بندی [۲۱] و غیره اشاره کرد. در این مقاله، ما از تبدیل موجک گسسته دو بعدی Haar استفاده کرده، و از تصاویر حاصل از این تبدیل به روشی جدید و به منظور یادگیری ویژگی‌های مهم با هدف بهبود مسأله‌ی تعمیم دامنه بهره می‌بریم.

۳ روش پیشنهادی

در مسأله‌ی تعمیم دامنه، هدف آموختن مدلی با استفاده از داده‌های آموزش دامنه‌های منبع است که انتظار می‌رود برای دامنه هدف نادیده که در مرحله‌ی آموزش در دسترس نبوده است، به خوبی عمل کند. کار قبلی که در این زمینه مطرح شد [۱۳] روشی به نام JiGen است که تأثیر حل پازل jigsaw در یادگیری نمایه عمومی تصاویر را نشان می‌دهد. شکل ۳، چارچوب کلی این روش را نشان می‌دهد. همان‌طور که در شکل مشاهده می‌شود، تصاویری از دامنه‌های گوناگون با برچسب‌های واقعی^۴ (برچسب‌های داده‌ای مربوط به کلاس آن‌ها) تحت عنوان تصاویر مرتب وجود دارند. این تصاویر با استفاده از تجزیه 3×3 به ۹ قطعه تبدیل می‌شوند، سپس با ترکیب تصادفی قطعه‌ها، تصاویری هم دامنه با تصاویر اصلی، تحت عنوان تصاویر بهم ریخته ایجاد می‌شوند. با توجه به ترتیب قطعه‌ها در تصاویر بهم ریخته، اندیسی^۵ تحت عنوان شبه برچسب برای آن‌ها تولید می‌شود. از کل ترتیب‌های موجود برای قطعه‌های یک تصویر، تنها p ترتیب با استفاده از الگوریتم مبتنی بر فاصله همینگ [۱۷] انتخاب می‌شود. تصاویر اصلی و تصاویر بهم ریخته بطور همزمان به یک شبکه عصبی پیچشی^۶ داده می‌شوند که برای برآورده کردن دو هدف بهینه می‌شود: طبقه بندی تصاویر اصلی (مرتب) با برچسب‌های واقعی و طبقه بندی تصاویر بهم ریخته با شبه برچسب‌های تولید شده برای آن‌ها. هدف دوم باعث می‌شود تا مدل، ویژگی‌های مربوط به همبستگی مکانی تصاویر را

بیاموزد، که می‌تواند برای مسأله‌ی تعمیم دامنه بسیار مفید باشد. برای این کار، از یک شبکه با دو لایه تماماً متصل^۷ متفاوتی که برای هر هدف به صورت مجزا (طبقه بندی تصاویر بهم ریخته و طبقه بندی تصاویر اصلی) در انتهای شبکه قرار دارد، استفاده شده است. در مرحله‌ی آزمون، طبقه بند آموخته شده با استفاده از تصاویر اصلی، بر روی داده‌های دامنه‌ی هدف اعمال می‌شود، تا عملکرد آن مورد ارزیابی قرار بگیرد. ما این روش را به عنوان اساس کار خود قرار داده و از چارچوبی مشابه با آن در روش پیشنهادی خود استفاده می‌کنیم، با این تفاوت که به جای حل پازل jigsaw و استفاده از تصاویر بهم ریخته در روش JiGen، از تصاویر حاصل از اعمال تبدیل موجک بر روی تصاویر اصلی، برای یادگیری نمایه عمومی تصاویر استفاده می‌کنیم.

در روش پیشنهادی ما، علاوه بر طبقه بندی تصاویر اصلی با برچسب‌های واقعی در یک حالت با نظارت، تصاویر حاصل از تبدیل موجک با شبه برچسب‌های تولید شده برای آنها نیز طبقه بندی می‌شوند. این کار به عنوان یک وظیفه‌ی خود - نظارتی می‌تواند باعث یادگیری ویژگی‌های عمومی و مفید در میان تصاویر دامنه‌های مختلف شود، که می‌تواند برای بهبود مسأله‌ی تعمیم دامنه بسیار کمک کننده باشد. در روش پیشنهادی ما نیز همانند روش JiGen، از یک شبکه عصبی پیچشی با دو لایه تماماً متصل متفاوتی که برای هر طبقه بند در انتهای شبکه قرار داده شده، استفاده می‌شود. در ادامه روش پیشنهادی خود را با جزئیات بیشتر و با فرمول شرح می‌دهیم.

فرض می‌کنیم S دامنه در مرحله آموزش در دسترس باشند، و هر دامنه شامل N_i نمونه برچسب دار باشد: $\{(x_j^i, y_j^i)\}_{j=1}^{N_i}$ به طوری که x_j^i ، زامین تصویر و $y_j^i \in \{1, \dots, C\}$ برچسب مربوط به کلاس آن را نشان می‌دهد. اولین هدف اساسی، یادگیری برچسب‌های واقعی داده از تصاویر اصلی در یک حالت با نظارت است. پس با فرض این که، x ورودی شبکه پیچشی و y برچسب واقعی آن باشد، هدف کمینه کردن تابع هزینه‌ای است که میزان خطای بین برچسب واقعی y و برچسب پیش بینی شده توسط مدل h با پارامترهای θ_f و θ_c را اندازه گیری کند. پارامتر θ_f مربوط به فضای ویژگی استخراج شده از شبکه پیچشی و پارامتر θ_c مربوط به لایه تماماً متصلی است که به عنوان طبقه بند نهایی در انتهای شبکه قرار دارد این تابع به صورت زیر تعریف می‌شود:

(۱)

$$\operatorname{argmin}_{\theta_f, \theta_c} \sum_{i=1}^S \sum_{j=1}^{N_i} l_c(h(x_j^i | \theta_f, \theta_c), y_j^i)$$

به عنوان دومین هدف از شبکه خواسته می‌شود که اندیس اختصاص داده شده به تصاویر حاصل از تبدیل موجک، که به عنوان شبه برچسب برای آنها تولید شده را پیش بینی کند. وقتی

¹ Image compressing

² Image denoising

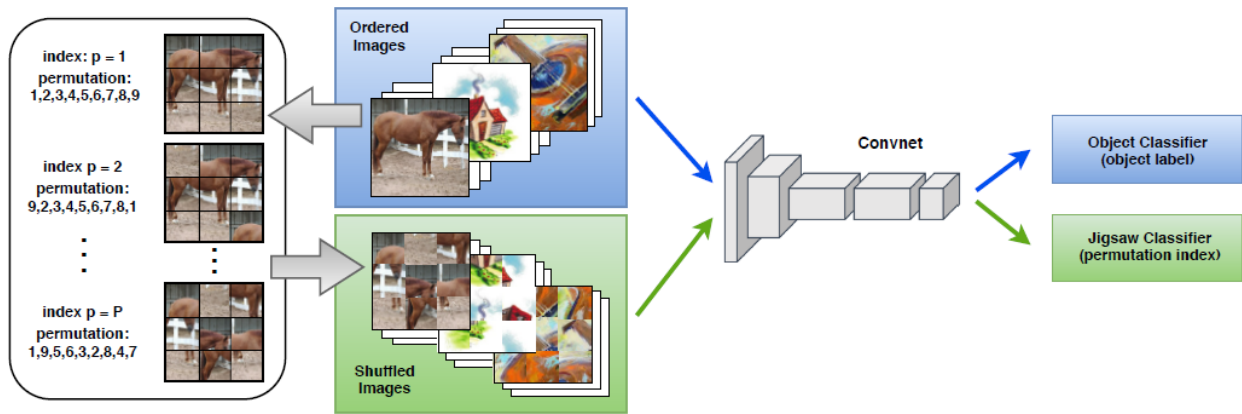
³ Image segmentation

⁴ True - lable

⁵ index

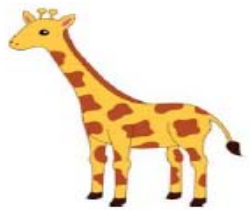
⁶ convolutional

⁷ Fully - connected layer

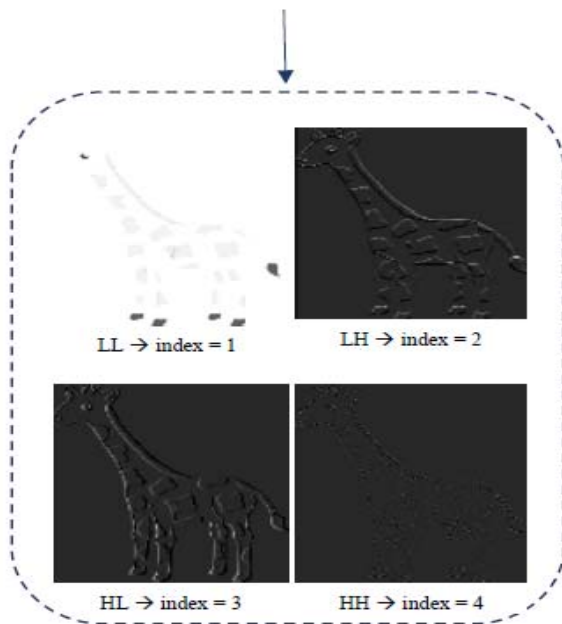


شکل ۳ چارچوب کلی روش JiGen [۱۳].

شبکه با توجه به این که تصویر وارد شده، تصویر اصلی یا یکی از تصاویر موجک است به لایه طبقه بندی مربوط به خود هدایت می-شوند. سپس، تابع هزینه‌ی مربوط به لایه طبقه بندی تصاویر اصلی با برجسب‌های واقعی و تابع هزینه‌ی مربوط به لایه طبقه بندی تصاویر موجک با شبه برجسب‌های تولید شده برای آنها، به طور مجزا محاسبه شده و با جای گذاری مقادیر به دست آمده در فرمول ۳، تابع هزینه نهایی محاسبه می‌شود. در نهایت عملیات پس انتشار^۲ و به روز رسانی پارامترهای شبکه انجام می‌گیرد.



Original image → index = 0



شکل ۴ زیرباندهای مختلف حاصل از تبدیل موجک گسسته دو بعدی.

که تبدیل موجک گسسته دو بعدی برای یک تصویر در یک سطح اعمال می‌شود، تصویر به چهار زیر باندها LL، LH، HL و HH تجزیه می‌شود. زیر باندها LL که زیرباند فرکانس پایین است مربوط به کلیات تصویر، زیرباندهای LH و HL که زیرباندهای میان فرکانسی هستند به ترتیب مربوط به جزئیات افقی و جزئیات عمودی، و زیرباندها HH که زیرباند فرکانس بالا است مربوط به جزئیات قطری می‌باشند. بنابراین همان طور که در شکل ۴ مشاهده می‌شود، به ازای هر تصویر چهار تصویر دیگر حاصل از تبدیل موجک گسسته داریم که به هر کدام یک اندیس به عنوان برجسب آن تصویر اختصاص می‌دهیم و یک وظیفه‌ی طبقه بندی دیگر تعریف می‌کنیم.

پس با فرض این که، z یکی از این K_i تصویر ایجاد شده به عنوان ورودی شبکه پیچشی و w برجسب تولید شده برای آن باشد، هدف کمینه کردن تابع هزینه زیر است:

(۲)

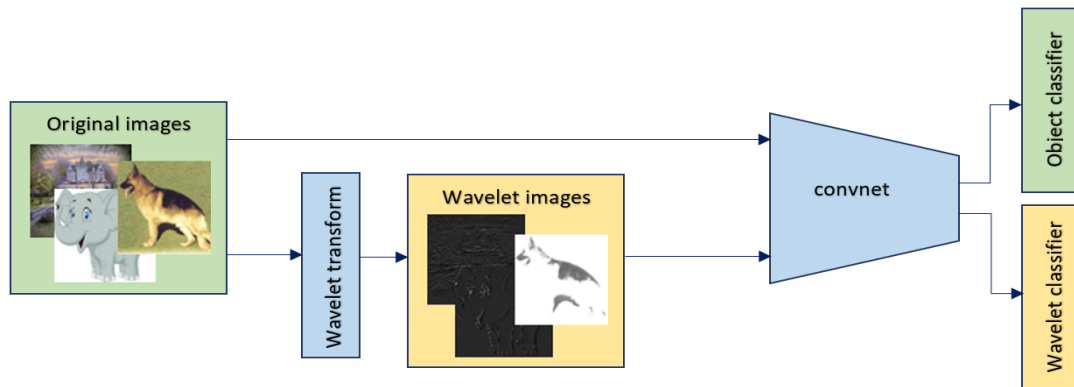
$$\operatorname{argmin}_{\theta_f, \theta_w} \sum_{i=1}^S \sum_{k=1}^{k_i} l_w(h(z_k^i | \theta_f, \theta_w), w_k^i)$$

با توجه به این که ساختار شبکه برای هر دو هدف یکسان و تنها در انتهای شبکه دو لایه تماماً متصل برای هر هدف به صورت مجزا در نظر گرفته شده، تابع مدل عمیق h یکسان و پارامتر θ_f مشترک است و پارامتر θ_w مربوط به لایه تماماً متصلی که برای پیش بینی اندیس تصاویر حاصل از تبدیل موجک اختصاص داده شده، می-باشد. در نهایت شبکه را با تابع نهایی زیر آموزش می‌دهیم:

$$\operatorname{argmin}_{\theta_f, \theta_c, \theta_w} \sum_{i=1}^S \sum_{j=1}^{N_i} l_c(h(x_j^i | \theta_f, \theta_c), y_j^i) + \sum_{k=1}^{k_i} \alpha l_w(h(z_k^i | \theta_f, \theta_w), w_k^i) \quad (3)$$

به طوری که توابع l_c و l_w از نوع بی نظمی متقابل^۱ هستند. شکل ۵، چارچوب کلی روش ما را نشان می‌دهد. همان طور که در شکل مشاهده می‌شود، تصاویر اصلی و تصاویر حاصل از تبدیل موجک به طور همزمان به شبکه وارد می‌شوند، و در انتهای

² Back-propagation¹ Cross entropy



شکل ۵ چارچوب کلی روش ارائه شده. تصاویر اصلی و تصاویر حاصل از تبدیل موجک همزمان به یک شبکه پیچشی داده می‌شوند و دو هدف مربوط به طبقه بندی تصاویر اصلی با برجسب‌های واقعی و طبقه بندی تصاویر حاصل از تبدیل موجک با شبه برجسب‌های تولید شده برای آنها بهینه می‌شوند.

ترتیب قطعه‌های تصویر بهم ریخته‌ای است که حاصل ترکیب تصادفی قطعه‌های حاصل از تجزیه $n \times n$ آن تصویر می‌باشد. از $n^2!$ ترتیب موجود برای قطعه‌های یک تصویر، تنها یک مجموعه‌ای با p عنصر با استفاده از الگوریتم مبتنی بر فاصله همینگ [۱۷] انتخاب می‌شود. در استفاده از این وظیفه، تصاویر اصلی به عنوان یک پازل مرتب همیشه در تابع هزینه jigsaw محاسبه می‌شوند اما تصاویر بهم ریخته تأثیری بر تابع هزینه طبقه بندی برجسب‌های واقعی ندارند. در وظیفه حدس زاویه چرخش، ما تصاویر را با استفاده از چهار زاویه ۰، ۹۰، ۱۸۰ و ۲۷۰ درجه به صورت تصادفی می‌چرخانیم و از شبکه می‌خواهیم که زاویه چرخش را حدس بزند. در این وظیفه تصویر اصلی که زاویه چرخش آن صفر درجه است همانند قبل در تابع هزینه چرخش محاسبه می‌شود اما تصاویر چرخیده تأثیری بر تابع هزینه طبقه بندی برجسب‌های واقعی ندارند.

۴ ارزیابی روش پیشنهادی

۴-۱ مجموعه داده

ما در این مقاله از مجموعه داده‌های معروف PACS [۱۰]، VLCS [۳۱] و Office-Home [۳۲] برای مسأله‌ی تعمیم دامنه استفاده می‌کنیم. مجموعه داده PACS دارای چهار دامنه، photo، art، painting، cartoon و sketch، و هفت گروه شیء، مجموعه داده VLCS دارای پنج گروه شیء و چهار دامنه مجزا، Sun، Caltech، LableMe، PASCALVOC2007، و مجموعه داده Office-Home دارای ۶۵ گروه شیء و چهار دامنه Product، Clipart، Art، Real-World می‌باشد. ما در هر بار آزمایش از یک دامنه به عنوان دامنه هدف برای آزمون و از سه دامنه دیگر به عنوان دامنه‌های منبع برای آموزش استفاده می‌کنیم و هر دامنه را به دو مجموعه برای آموزش (۷۰ درصد) و آزمون (۳۰ درصد) با انتخاب تصادفی روی کل دامنه، تقسیم می‌کنیم.

توجه داشته باشید که، در زمان آموزش، تابع هزینه موجک هم برای تصاویر اصلی و هم برای تصاویر موجک، محاسبه می‌شود (یعنی تصاویر اصلی با اندیس صفر همیشه وارد لایه طبقه بندی تصاویر موجک می‌شوند و تابع هزینه موجک برای آنها محاسبه می‌شود). این در حالی است که تابع هزینه طبقه بندی برجسب‌های واقعی، فقط تحت تأثیر تصاویر اصلی قرار می‌گیرد. در زمان آزمون نیز ما تنها از لایه مربوط به طبقه بندی تصاویر اصلی استفاده می‌کنیم تا برجسب تصاویر دامنه‌ی هدف جدید را پیش بینی کند.

در این روش، به طور کلی با دو پارامتر α و β سروکار داریم. پارامتر α که در فرمول ۳ مشاهده می‌شود، وزن تابع هزینه مربوط به طبقه بندی تصاویر موجک است که میزان تأثیر این تابع در تابع هزینه نهایی را تعیین می‌کند و مقداری بین ۰ و ۱ دارد. پارامتر β که پارامتر گرایش داده^۱ نامیده می‌شود، نیز مقداری بین ۰ و ۱ دارد و تعیین می‌کند در هر دسته^۲ ای از تصاویر که وارد شبکه می‌شوند چند درصد از آنها تصاویر اصلی و چند درصد تصاویر موجک باشند. به طور مثال، اگر β برابر ۰٫۶ باشد یعنی برای هر دسته ۶۰٪ تصاویر اصلی و ۴۰٪ تصاویر موجک هستند.

زمانی که بخواهیم از وظایف حل پازل jigsaw، حدس زاویه چرخش و تبدیل موجک دو به دو یا هر سه به صورت همزمان استفاده کنیم، کافی است تابع هزینه بعدی را با ضریب متفاوتی در فرمول نهایی تکرار کنیم. به طور مثال، برای زمانی که بخواهیم از حل پازل jigsaw و تبدیل موجک استفاده کنیم تابع هزینه نهایی به شکل زیر محاسبه می‌شود:

$$L_c + \alpha L_w + \gamma L_j \quad (4)$$

به طوری که L_j تابع هزینه مربوط به تصاویر بهم ریخته و γ ضریب مربوط به آن است. در وظیفه حل پازل jigsaw همان طور که در [۱۳] به طور کامل تشریح شده، هدف پیش بینی

¹ Data - bias

² Batch

۴-۲ جزئیات پیاده سازی

همان طور که قبل تر گفته شد، ما در این مقاله برای انجام آزمایش ها از تبدیل موجک گسسته دو بعدی Haar در یک سطح استفاده کردیم. از آنجایی که برای اعمال این تبدیل بر روی یک تصویر، بایستی آن تصویر در مقیاس خاکستری^۱ باشد، ابتدا آن را به مقیاس خاکستری تبدیل کرده، سپس تبدیل موجک را بر روی آن اعمال کردیم. تصاویر حاصل از تبدیل موجک نیز خاکستری بوده و دارای اندازه‌های کوچکتر از تصویر اصلی هستند. با توجه به این که تصاویر اصلی، رنگی و در شبکه پیچشی دارای سه کانال R, G و B هستند و به دلیل اینکه تصاویر اصلی و تصاویر موجک با هم وارد شبکه می‌شوند بنابراین، بعد از اعمال تبدیل موجک، تصاویر حاصل را به حالت RGB و به اندازه تصاویر اصلی تبدیل کردیم تا تمام تصاویر ورودی به شبکه، دارای شرایط یکسانی باشند. برای تبدیل تصاویر به حالت RGB، از یک عملیات ساده پردازش تصویر به نام "convert" از یک کتابخانه‌ی معروف پایتون به نام "PIL" استفاده کردیم که تصاویر خاکستری تک کانال را به سه کانال رنگی تبدیل می‌کند. علاوه بر این، به منظور تأثیر بیشتر ویژگی‌های استخراج شده، تصاویر حاصل از تبدیل موجک را با یک عدد ثابت جمع کردیم.

همان طور که گفته شد، در این روش دو پارامتر α و β داریم که α وزن تابع هزینه موجک و β میزان تمایل به تصاویر اصلی و تصاویر موجک را نشان می‌دهد. به منظور مقایسه‌ی عادلانه‌ی تبدیل موجک با پازل jigsaw از تنظیمات ارائه شده برای روش JiGen [۱۳] استفاده کردیم. بنابراین، دو شبکه resnet (با $\alpha=0,9$ و $\beta=0,9$) و alexnet (با $\alpha=0,9$ و $\beta=0,7$) را برای آموزش مدل، و تعداد دورها^۲ را برابر با ۳۰ در نظر گرفتیم تا شرایط کاملاً یکسان باشد. در همه‌ی آزمایش‌ها، ما شبکه را با استفاده از بهینه ساز^۳ SGD با اندازه دسته^۴ ۱۲۸ و نرخ یادگیری^۵ ۰,۰۰۱ (که بعد از ۸۰٪ دورها به ۰,۰۰۱ کاهش می‌یابد) آموزش دادیم. همچنین، از برش تصادفی^۶، چرخش افقی تصادفی^۷، تبدیل تصادفی به مقیاس خاکستری و غیره نیز برای افزایش داده‌ها^۸ استفاده کردیم. در آزمایش‌هایی که از ترکیب وظایف خود - نظارتی استفاده شده، تنها شبکه resnet را برای ۵۰ دور با پارامتر β برابر ۰,۹ آموزش داده و نتایج را مقایسه کردیم.

۴-۳ نتایج

در این قسمت، ما نتایج آزمایش‌های انجام شده را ارائه کرده و به تحلیل آنها می‌پردازیم. در ابتدا، روش پیشنهادی خود یعنی تبدیل موجک را با پازل jigsaw در روش JiGen به عنوان نزدیک ترین روش و چندین روش تعمیم دامنه اخیر مقایسه کرده و سپس نشان می‌دهیم که ترکیب دو یا چند وظیفه می‌تواند نتایج بسیار بهتری داشته باشد.

MMLD [۳۳]، روشی است که با استفاده از ترکیبی از چندین دامنه نهفته (بدون برچسب دامنه به عنوان مثال: Art, Photo, sketch و ...)، مدلی را برای تعمیم دامنه آموزش می‌دهد. در این روش با خوشه بندی ویژگی‌های استخراج شده از لایه‌های پایین تر، شبه برچسب‌های دامنه تولید شده و از طریق یادگیری خصمانه، ویژگی‌های ثابت در میان دامنه‌ها را یاد می‌گیرد. MLDG [۳۴] یک روش مبتنی بر فرا یادگیری است که تغییر دامنه آموزش/آزمون را در زمان آموزش شبیه سازی می‌کند و S-MLDG [۳۵]، یک چارچوب مبتنی بر یادگیری متوالی^۹ را برای این روش ارائه می‌دهد. SNR [۳۶]، روشی است که یک ماژول نرمال سازی و بازگردانی^{۱۰} سبک^{۱۱}، به هدف یادگیری نمایه ویژگی‌های تعمیم پذیر در میان دامنه‌های مختلف را برای تعمیم دامنه و تطبیق دامنه مؤثر ارائه می‌دهد.

جدول ۱، دقت مدل را با استفاده از مجموعه داده‌ی PACS نشان می‌دهد.

جدول ۱ دقت تعمیم دامنه برای مجموعه داده‌ی PACS با استفاده از شبکه‌های alexnet و resnet 18. کلیه دقت‌ها، حاصل میانگین سه مرتبه اجرا هستند. هر ستون نام دامنه ای است که به عنوان دامنه هدف در نظر گرفته شده است. ما از فونت پررنگ برای جاهایی که دقت ما بهتر بوده و از فونت زیرخط برای جاهایی که دقت سایر روش‌ها بهتر است، استفاده کردیم.

Methods	PACS dataset				Avg.
	art paint.	cartoon	sketches	photo	
Alexnet					
JiGen	۶۷,۶۳	۷۱,۷۱	۶۵,۱۸	۸۹,۰۰	۷۳,۳۸
MMLD	۶۹,۲۷	۷۲,۸۳	۶۶,۴۴	۸۸,۹۸	۷۴,۳۸
Ours	۶۶,۹۳	۷۱,۱۱	۷۰,۴۴	۹۰,۶۴	۷۴,۷۸
Resnet 18					
JiGen	۷۹,۴۲	۷۵,۲۵	۷۱,۳۵	۹۶,۰۳	۸۰,۵۱
S-MLDG	۸۰,۵۰	۷۷,۸۰	۷۲,۸۰	۹۴,۸۰	۸۱,۵۰
SNR	۸۰,۳۰	۷۸,۲۰	۷۴,۱۰	۹۴,۵۰	۸۱,۸۰
MMLD	۸۱,۲۸	۷۷,۱۶	۷۲,۲۹	۹۶,۰۹	۸۱,۸۳
Ours	۸۰,۶۹	۷۵,۲۲	۷۶,۶۵	۹۶,۵۳	۸۲,۲۷

¹ Gray scale

² epochs

³ Optimizer

⁴ Batch - size

⁵ Learning rate

⁶ Random crop

⁷ Random horizontal flip

⁸ Data augmentation

⁹ sequential

¹⁰ restitute

¹¹ style

همان طور که در شکل مشخص است روش پیشنهادی ما می-تواند بهتر از روش JiGen عمل کند و به مدل کمک می کند تا نمایه عمومی در میان دامنه های مختلف را به خوبی بیاموزد. ما روش پیشنهادی خود را با استفاده از مجموعه داده های VLCS و Office-Home نیز مورد آزمایش قرار داده و دقت مدل را به ترتیب در جداول ۲ و ۳ ارائه دادیم. همان طور که مشاهده می-شود، روش ما با استفاده از این مجموعه داده ها نیز دارای عملکرد خوبی بوده و می تواند به طور میانگین از روش های دیگر بهتر عمل کند.

جدول ۲ دقت تعمیم دامنه برای مجموعه داده ی VLCS با استفاده از شبکه ی alexnet. تعداد اجراها، معنی ستون ها و استفاده از فونت پررنگ یا زیر خط همانند جدول ۱ می باشد.

Methods	VLCS dataset				
	Caltech	Labelme	Pascal	Sun	Avg.
	Alexnet				
JiGen	۹۶,۹۳	۶۰,۹۰	۷۰,۶۲	۶۴,۳۰	۷۳,۱۹
S-MLDG	۹۶,۴۰	۶۴,۸۰	۶۸,۷۰	۶۴,۰۰	۷۳,۵۰
MMLD	۹۶,۶۶	۵۸,۷۷	۷۱,۹۶	۶۸,۱۳	۷۳,۸۸
Ours	۹۷,۶۴	۶۲,۰۲	۷۲,۲۳	۶۵,۶۵	۷۴,۳۸

جدول ۳ دقت تعمیم دامنه برای مجموعه داده ی Office-Home با استفاده از شبکه ی resnet 18. تعداد اجراها، معنی ستون ها و استفاده از فونت پررنگ یا زیر خط همانند جدول ۱ می باشد.

Methods	Office-Home dataset				Avg.
	Art	Clipart	Product	Real-World	
	Resnet 18				
JiGen	۵۳,۰۴	۴۷,۵۱	۷۱,۴۷	۷۲,۷۹	۶۱,۲۰
Ours	۵۵,۶۹	۴۸,۳۴	۷۰,۹۹	۷۵,۳۰	۶۲,۵۸

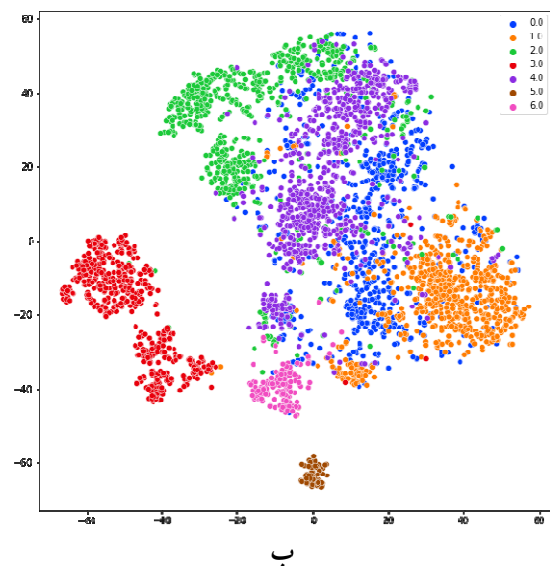
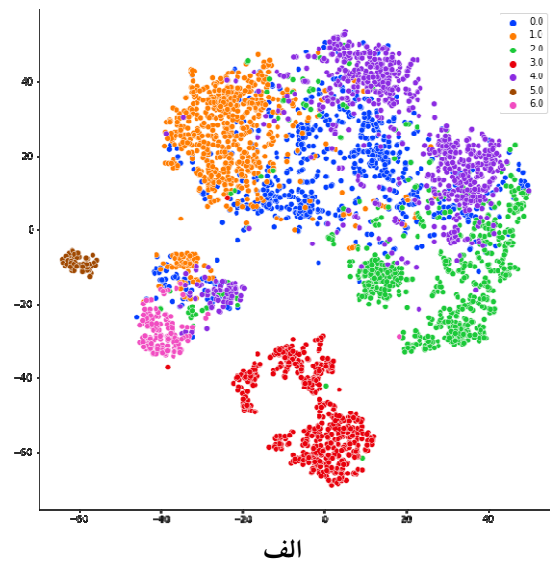
همان طور که در [۱۳] ذکر شده است، دقت به دست آمده برای وظیفه ی حدس زاویه چرخش با پروتکل JiGen نسبت به پازل jigsaw کمتر است بنابراین از آوردن نتایج آن به تنهایی صرف نظر کردیم. اما در ادامه چند وظیفه ی پازل jigsaw، حدس زاویه چرخش و روش پیشنهادی ما یعنی تبدیل موجک را با هم و به صورت دو به دو ترکیب کردیم.

با توجه به آن چه که در جدول ۴ آمده است، می بینیم که ترکیب چند وظیفه با هم می تواند منجر به یادگیری بهتر ویژگی های عمومی شده و دقت بهتری را برای مسأله ی تعمیم دامنه رقم بزند. همان طور که در این جدول نیز مشخص است، تبدیل موجک به شدت باعث بهبود دامنه ی sketches شده است و این موضوع می تواند برای تحقیقات آینده بسیار مورد توجه قرار گیرد.

تأثیر پارامترها: همان طور که قبل تر گفته شد، ما در این روش با دو پارامتر α و β سرو کار داریم. در این بخش با ثابت نگه داشتن یک پارامتر و تغییر پارامتر دیگر، تأثیر پارامترها روی مدل را مورد بررسی قرار داده و نتایج را در شکل ۷ ارائه دادیم. در قسمت الف، پارامتر β را برابر ۰,۹ قرار داده (یعنی ۹۰ درصد

همان طور که در جدول می بینیم، روش پیشنهادی ما در دامنه sketches با تفاوت زیادی نسبت به سایر روش ها بهتر عمل می-کند. چرا که این دامنه، طرح اولیه یک شیء را به تصویر کشیده و ساختار کلی آن را حفظ می کند. علاوه بر این، در دامنه sketches تصاویر، بافت خاصی ندارند و همین خصوصیات باعث شده که مدل آموخته شده با دامنه های دیگر، بتواند به خوبی برای این دامنه خاص تعمیم یافته و دارای عملکرد بهتری باشد. این در حالی است که روش ما در سایر دامنه ها می تواند به خوبی سایر روش ها و حتی در مواردی بهتر عمل کند. به همین دلیل، روش پیشنهادی ما به طور میانگین بهتر از روش JiGen و سایر روش های ذکر شده است.

به منظور مقایسه ی بیشتر روش پیشنهادی با روش مشابه، نمودار t-SNE مربوط به روش JiGen در قسمت الف و روش پیشنهادی ما در قسمت ب از شکل ۶ برای دامنه هدف sketches رسم شده است.

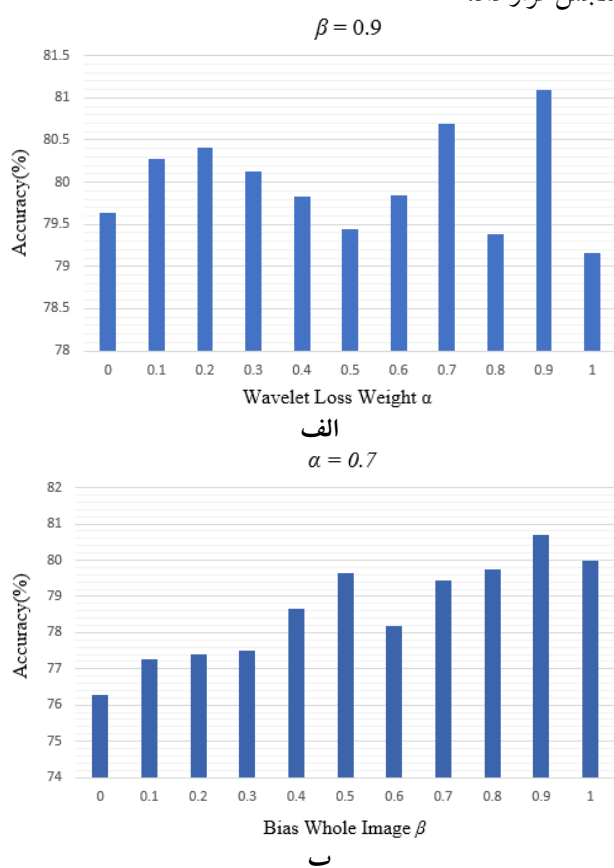


شکل ۶ نمودار t-SNE برای دامنه sketches به عنوان دامنه هدف از مجموعه داده PACS با شبکه resnet 18. قسمت الف مربوط به روش JiGen و قسمت ب مربوط به روش پیشنهادی ما است.

جدول ۴ دقت تعمیم دامنه با ترکیبی از وظایف خود-نظارتی برای مجموعه داده PACS با استفاده از شبکه‌ی 18 resnet. سطر آخر ترکیب هر سه وظیفه پازل jigsaw، حدس زاویه چرخش تصویر و تبدیل موجک است. کلیه دقت‌ها، حاصل میانگین سه مرتبه اجرا هستند. هر ستون نام دامنه ای است که به عنوان دامنه هدف در نظر گرفته شده است. در همه ی آزمایش‌ها $W_r=0.8$ و $W_w=0.6$ ، $W_j=0.6$ در نظر گرفته شده است، که اولی وزن پازل jigsaw، دومی وزن تبدیل موجک و سومی وزن حدس زاویه چرخش در تابع هزینه می باشند. ما از فونت پرننگ برای جاهایی استفاده کردیم که بهترین دقت حاصل شده است.

Methods	PACS dataset				Avg.
	art paint.	cartoon	sketches	Photo	
jigsaw & rotation	۸۱,۵۴	۷۷,۶۳	۷۴,۷۲	۹۶,۳۵	۸۲,۵۶
jigsaw & wavelet	۸۲,۳۰	۷۵,۵۰	۷۸,۲۱	۹۶,۸۱	۸۳,۲۰
wavelet & rotation	۸۱,۷۷	۷۸,۵۸	۷۷,۳۲	۹۶,۰۳	۸۳,۴۲
wavelet & jigsaw & rotation	۸۲,۵۱	۷۸,۴۴	۷۷,۰۷	۹۵,۹۷	۸۳,۴۹

- نظارتی دیگر همچون حل پازل jigsaw و حدس زاویه چرخش با یکدیگر و با روش پیشنهادی ما یعنی تبدیل موجک می‌تواند نتایج بسیار بهتری داشته باشد. استفاده از تبدیل موجک گسسته به صورت یک وظیفه‌ی خود - نظارتی کار جدیدی است که در این مقاله به آن پرداخته شده و می‌تواند به عنوان یک موضوع تحقیقاتی جدید برای آینده مورد استفاده قرار گیرد، همچنین با توجه به سادگی این روش، می‌توان آن را برای کاربردهای دیگری از جمله قطعه بندی تصاویر نیز مورد سنجش قرار داد.



شکل ۷ بررسی تأثیر پارامترهای α و β روی دقت مدل با استفاده از دامنه هدف art_painting از مجموعه داده PACS و شبکه resnet-18. کلیه دقت‌ها حاصل میانگین سه مرتبه اجرا هستند.

تصاویر موجود در هر دسته، تصاویر اصلی و تنها ۱۰ درصد آنها تصاویر موجک هستند) و با تغییر پارامتر α در بازه‌ی ۰ تا ۱، تأثیر این پارامتر را روی دقت مدل بررسی کردیم. زمانی که α برابر ۰ و β برابر ۱ باشد، یعنی وظیفه‌ی طبقه بندی تصاویر موجک خاموش است زیرا تنها تصاویر اصلی وارد شبکه شده و تابع هزینه موجک تأثیری در تابع هزینه نهایی ندارد. همان طور که در شکل قسمت الف مشاهده می‌شود، می‌توان گفت دقت در همه حالت‌ها نسبت به زمانی که $\alpha=0$ و $\beta \sim 1$ است، تقریباً بهتر یا مساوی بوده و زمانی که مقدار α برابر ۰,۹ است و تابع هزینه موجک تأثیر زیادی در تابع هزینه نهایی دارد، دقت در بالاترین مقدار خود قرار دارد. از طرفی فاصله بین کمترین دقت و بیشترین دقت تنها ۲ درصد است و این یعنی پارامتر α باعث تغییرات زیادی در نتایج نمی‌شود و مدل به این پارامتر حساسیت زیادی ندارد. در قسمت ب، پارامتر α را برابر ۰,۷ قرار داده و با تغییر پارامتر β در بازه ۰ تا ۱، تأثیر این پارامتر را نیز مورد بررسی قرار دادیم. زمانی که $\beta < 0,5$ باشد، تمرکز روی طبقه بندی تصاویر موجک و زمانی که $\beta > 0,5$ ، تمرکز روی طبقه بندی تصاویر اصلی است. همان طور که در شکل قسمت ب، مشاهده می‌شود با افزایش این پارامتر دقت نیز بهتر می‌شود زیرا طبقه بندی تصاویر اصلی به عنوان هدف اصلی در یادگیری مدل، انتخاب می‌شود.

۵ جمع بندی

در این مقاله، ما نشان دادیم که یادگیری همزمان برچسب‌های داده از تصاویر اصلی به منظور طبقه بندی و یادگیری ویژگی‌های عمومی با استفاده از تبدیل موجک در یک حالت خود - نظارت می‌تواند به طور مؤثری به مسئله‌ی تعمیم دامنه کمک کند. ما شبکه پیچشی عمیق را مجبور کردیم تا تصاویری که از تصویر اصلی با اعمال تبدیل موجک گسسته به دست می‌آیند را با شبه برچسب‌هایی که برای آنها تولید شده، طبقه بندی کند. این کار باعث یادگیری ویژگی‌هایی می‌شود که می‌تواند به مسئله‌ی تعمیم دامنه بسیار کمک کند. ما نیز نشان دادیم که ترکیب وظایف خود

مراجع

- [12] Mohammad Mahfujur Rahman. Clinton Fookes. Mahsa Baktashmotlagh. Sridha Sridharan. "Multi-component Image Translation for Deep Domain Generalization". IEEE Winter Conference on Applications of Computer Vision (WACV). 2019.
- [13] Fabio M Carlucci. Antonio D'Innocente. Silvia Bucci. Barbara Caputo. and Tatiana Tommasi. "Domain generalization by solving jigsaw puzzles". In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pages 2229–2238. 2019.
- [14] Mei Wang. Weihong Deng. "Deep Visual Domain Adaptation: A Survey". Neurocomputing. Volume 312. Pages 135–153. 2018.
- [15] Ya Li. Xinmei Tian. Mingming Gong. Yajing Liu. Tongliang Liu. Kun Zhang. and Dacheng Tao. "Deep domain generalization via conditional invariant adversarial networks". In The European Conference on Computer Vision (ECCV). pages 624–639. 2018.
- [16] Longlong Jing and Yingli Tian. "Self-supervised visual feature learning with deep neural networks: A survey". in IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020.
- [17] Mehdi Noroozi and Paolo Favaro. "Unsupervised learning of visual representations by solving jigsaw puzzles". In European Conference on Computer vision. pages 69–84. Springer. 2016.
- [18] Dahun Kim. Donghyeon Cho. Donggeun Yoo. and In So Kweon. "Learning image representations by completing damaged jigsaw puzzles". In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). pages 793–802. IEEE. 2018.
- [19] Spyros Gidaris. Praveer Singh. and Nikos Komodakis. "Unsupervised representation learning by predicting image rotations". arXiv abs/1803.07728. 2018.
- [20] Yu Sun. Eric Tzeng. Trevor Darrell. Alexei A. Efros. "Unsupervised Domain Adaptation Through Self-Supervision". eprint arXiv:1909.11825. 2019.
- [21] Rasheed A. Younis M.S.. Bilal M.. Rasheed M. "Classification of Chest Diseases Using Wavelet Transforms and Transfer Learning". In Medical Imaging and Computer-Aided Diagnosis(MICAD). Lecture Notes in Electrical Engineering. vol 633. Springer. Singapore. 2020.
- [22] Haixia Bi. Lin Xu. Xiangyong Cao. Yong Xue. and Zongben Xu. "Polarimetric SAR Image Semantic Segmentation with 3D Discrete Wavelet Transform and Markov Random Field". in IEEE Transactions on Image Processing. vol. 29. pp. 6601–6614. 2020.
- [23] Garima Singh. Pushpa Koranga. Dikendra Verma. "Application of Wavelet Transform on Images: A Review". International Journal of Engineering Research and General Science. Volume 4. Issue 5. September–October. 2016.
- [1] G. Blanchard. G. Lee. and C. Scott. "Generalizing from several related classification tasks to a new unlabeled sample". Proceedings of the 24th International Conference on Neural Information Processing Systems(NIPS).Pages 2178–2186. 2011.
- [2] K. Muandet. D. Balduzzi. and B. Schölkopf. "Domain generalization via invariant feature representation". Proceedings of the 30th International Conference on Machine Learning(ICML). PMLR 28(1):10–18. 2013.
- [3] Z. Xu. W. Li. L. Niu. and D. Xu. "Exploiting low-rank structure from latent domains for domain generalization". In European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science. vol 8691. Springer. Cham. 2014.
- [4] M. Ghifary. W. B. Kleijn. M. Zhang. and D. Balduzzi. "Domain generalization for object recognition with multi-task autoencoders". IEEE International Conference on Computer Vision (ICCV). pp. 2551–2559. 2015.
- [5] Khosla. T. Zhou. T. Malisiewicz. A. A. Efros. and A. Torralba. "Undoing the damage of dataset bias". In Proceedings of the European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science. vol 7572. Springer. Berlin. Heidelberg. 2012.
- [6] Fang. Y. Xu. and D. N. Rockmore. "Unbiased metric learning: On the utilization of multiple datasets and web images for softening bias". In IEEE International Conference on Computer Vision (ICCV). pp. 1657–1664. 2013.
- [7] L. Niu. W. Li. and D. Xu. "Multi-view domain generalization for visual recognition". In IEEE International Conference on Computer Vision (ICCV). pp. 4193–4201. 2015.
- [8] L. Niu. W. Li. D. Xu. and J. Cai. "An exemplar-based multiview domain generalization framework for visual recognition". in IEEE Transactions on Neural Networks and Learning Systems. vol. 29. no. 2. pp. 259–272. 2018.
- [9] S. Motiian. M. Piccirilli. D. A. Adjero. and G. Doretto. "Unified deep supervised domain adaptation and generalization". Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 5715–5725. 2017.
- [10] D. Li. Y. Yang. Y. Z. Song. and T. M. Hospedales. "Deeper. broader and artier domain generalization". In IEEE International Conference on Computer Vision (ICCV). pages 5542–5550. 2017.
- [11] Haoliang Li. Sinno Jialin Pan. Shiqi Wang. and Alex C. Kot. "Domain generalization with adversarial feature learning". In Computer Vision and Pattern Recognition (CVPR). pages 5400–5409. 2018.



سارا فرهمندی نیا مدرک کارشناسی خود را در سال ۱۳۹۵ در رشته مهندسی کامپیوتر گرایش نرم افزار با کسب رتبه اول از دانشگاه شهید باهنر کرمان اخذ نمود. ایشان هم اکنون دانشجوی کارشناسی ارشد در رشته مهندسی کامپیوتر گرایش هوش مصنوعی در همان دانشگاه می باشد. حوزه های پژوهش مورد علاقه ایشان، یادگیری عمیق و پردازش تصویر می باشند.



مهدی افتخاری مدرک کارشناسی خود را در سال ۱۳۷۹ در رشته مهندسی کامپیوتر گرایش سخت افزار از دانشگاه شیراز و مدرک کارشناسی ارشد و دکترای خود را به ترتیب در سال های ۱۳۸۲ و ۱۳۸۶ در رشته مهندسی کامپیوتر گرایش هوش مصنوعی از همان دانشگاه اخذ کرد. وی از سال ۱۳۸۶ تاکنون عضو هیأت علمی بخش مهندسی کامپیوتر، دانشکده فنی و مهندسی، دانشگاه شهید باهنر کرمان است و در سال ۱۳۹۳ به مرتبه دانشیاری ارتقاء پیدا کرده است. حوزه های تخصصی پژوهش ایشان یادگیری ماشین، یادگیری عمیق و مجموعه ها و سیستم های فازی است. وی تاکنون بیش از ۱۳۰ مقاله علمی در نشریات و کنفرانس های معتبر داخلی و خارجی به چاپ رسانیده است.



کاوه بهرامن در حال حاضر دانشجوی سال سوم کارشناسی مهندسی کامپیوتر دانشگاه شهید باهنر کرمان است. حوزه های پژوهش مورد علاقه ایشان یادگیری ماشین، یادگیری عمیق و ریاضیات می باشند.

- [24] Seyede Mahya Hazavei, Hamid Reza Shahdoosti. "Using Complex Wavelet Transform and Bilateral Filtering for Image Denoising". eprint arXiv:1702.01276. 2017.
- [25] Tom Mitchell, Bruce Buchanan, Gerald DeJong, Thomas Dietterich, Paul Rosenbloom, and Alex Waibel. "Machine learning". Annual review of computer science, 4(1):417-433, 1990.
- [26] Shoubo Hu, Kun Zhang, Zhitang Chen, and Laiwan Chan. "Domain generalization via multidomain discriminant analysis". Uncertainty in artificial intelligence : proceedings of the ... conference. Conference on Uncertainty in Artificial Intelligence, 35, 2019.
- [27] Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin. "Generalizing to Unseen Domains: A Survey on Domain Generalization". Preprint arXiv: 2103.03097v2, 2021.
- [28] M. Mancini, S. R. Bulò, B. Caputo, and E. Ricci. "Best sources forward: domain generalization through source-specific nets". In ICIP, pp. 1353-1357, 2018.
- [29] D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales. "Learning to generalize: Meta-learning for domain generalization". in AAAI, 2018.
- [30] Y. Balaji, S. Sankaranarayanan, and R. Chellappa. "Metareg: Towards domain generalization using meta-regularization". in NeurIPS, pp. 998-1008, 2018.
- [31] Antonio Torralba and Alexei A. Efros. "Unbiased look at dataset bias". In Conference on Computer Vision and Pattern Recognition (CVPR), 2011.
- [32] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. "Deep hashing network for unsupervised domain adaptation". In Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [33] T. Matsuura and T. Harada. "Domain generalization using a mixture of multiple latent domains". In AAAI, 2020.
- [34] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M. Hospedales. "Learning to generalize: Meta-learning for domain generalization". In Conference of the Association for the Advancement of Artificial Intelligence (AAAI), 2018.
- [35] Li D., Yang Y., Song YZ., Hospedales T. "Sequential Learning for Domain Generalization". ECCV, 2020.
- [36] Xin Jin, Cuiling Lan, Wenjun Zeng, Zhibo Chen. "Style Normalization and Restitution for Domain Generalization and Adaptation", arXiv:2101.00588, 2021.