

می‌شود. اگر این نگاشت را طبق نمادگذاری معرفی شده با ϕ نشان دهیم دسته بندی با استفاده از رابطه زیر انجام خواهد شد:

$$y^* = \arg \max_y \phi(c^y)^T x \quad (24-2)$$

برای یادگیری $\phi(c)$ از ترکیب دو تخمین‌گر استفاده می‌شود:

۱. رگرسیون احتمالی: توزیع P_{reg} یادگرفته می‌شود که برای یک توصیف c و نگاشت در فضای تصاویر w احتمال $P_{reg}(w|c)$ را مدل می‌کند.

۲. تابع مطابقت: نگاشت دو خطی D که تطابق میان دامنه تصاویر و توصیف‌ها مدل می‌کند به عبارت دیگر $c^T D x$ زمانی که x به دسته‌ای که c توصیف می‌کند تعلق دارد بزرگتر از مقدار آستانه‌ای است و در غیر این صورت کوچک‌تر از آن. می‌توان مشاهده کرد که در این حالت با استفاده از رابطه (24-2)، $c^T W$ یک دسته‌بند خطی برای دسته‌ای که c توصیف می‌کند، خواهد بود.

پارامترهای P_{reg} و D با استفاده از نمونه‌های آموزش بدست می‌آیند. در نهایت تابع پیشنهادی برای نگاشت ϕ برای دسته‌های آزمون به صورت زیر تعریف می‌شود:

$$\phi(c) = \arg \min_{w, \zeta_i} w^T w - \alpha c^T D w - \beta \ln(P_{reg}(w|c)) + \gamma \sum \zeta_i \quad (25-2)$$

$$s.t. : -(w^T x_i) \geq \zeta_i, \quad \zeta_i \geq 0, \quad i = 1, \dots, N_s$$

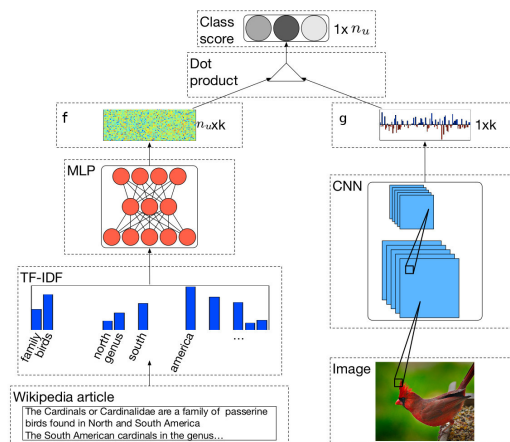
$$c^T D c \geq l$$

که α, β, γ, l فرایارامترهای مدل هستند. جمله اول در این تابع هزینه، منظم‌سازی دسته‌بند خطی w است. جمله دوم مشابهت w با $c^T D$ را الزام می‌کند و جمله سوم، مقدار راستی‌نمایی یک رگرسیون احتمالی برای تخمین w از x است. محدودیت $-(w^T x_i) \geq \zeta_i$ بر اساس فرض عدم تعلق نمونه‌های آزمون به کلاس‌های دیده‌شده تعریف شده است و اجبار می‌کند که تمامی نمونه‌های دیده‌شده باید در طرف منفی دسته‌بند خطی w قرار گیرند. نویسندگان این پژوهش، روش خود را با استفاده از تکنیک هسته ^{۲۸} برای دسته‌بندهای غیرخطی نیز توسعه داده‌اند [۳۶].

^{۲۸}kernel trick

۷-۲ نگاشت به یک فضای میانی

در برخی روش‌ها هر دوی نگاشت‌های ϕ و θ ، معرفی شده در ابتدای فصل با توجه به داده‌ها یاد گرفته می‌شوند و در نتیجه فضای مشترک مورد استفاده نه فضای تصاویر و نه فضای توصیف‌هاست؛ بلکه فضای ثالثی است. این فضای میانی در برخی از روش‌ها یک فضای با بعد کمتر است و تعبیر معنایی برای آن موجود نیست. در برخی روش‌های دیگر، فضای میانی را با بعد n_s یعنی تعداد دسته‌های دیده شده در نظر گرفته‌اند و تعبیر معنایی برای آن ارائه شده است. این فضای میانی بر اساس توصیف دسته‌ها و نمونه‌های دیده نشده بر اساس شباهت آن‌ها با دسته‌های دیده شده استوار است.



شکل ۲-۲: شبکه مورد استفاده برای یادگیری توام نگاشت تصاویر و توصیف‌ها که یک شبکه عصبی عمیق با دو ورودی است. ورودی اول از نوع تصویر است و ابتدا با یک شبکه پیچشی سپس با چند لایه چگال به فضایی k -بعدی می‌رود. ورودی دوم که یک مقاله از ویکی‌پدیای انگلیسی است پس از تبدیل به نمایش برداری به صورت $tf-idf$ با چند لایه با اتصالات چگال پردازش شده و به فضایی k -بعدی می‌رود. در نهایت امتیاز تعلق تصویر به دسته‌ی متن با ضرب داخلی این دو نگاشت تعیین می‌شود [۱].

در [۱] از شبکه‌های عصبی عمیق برای یادگیری توام نگاشت‌های ϕ و θ استفاده شده است. نمای کلی شبکه مورد استفاده در این روش در تصویر ۲-۲ نشان داده شده است. توصیف‌های متنی و ویژگی‌های بصری دو ورودی جداگانه به چنین شبکه‌ای هستند که ابتدا به صورت جداگانه با یک یا چند لایه‌ی با اتصالات کامل به یک فضای مشترک نگاشته شده و سپس بر اساس شباهت نمایش آن‌ها در این فضای میانی دسته‌بندی می‌شوند. تفاوت این روش با سایر روش‌هایی که مرور شد یادگیری توامان نگاشت‌های ϕ و θ است که با استفاده از شبکه‌های عصبی ممکن شده است. معیار یادگیری این دو نگاشت تنها خطای دسته‌بندی نهایی است. این روش را می‌توان به صورت ساخت دسته‌بند از روی توصیفات نیز تعبیر کرد؛ با این تفاوت که در این حالت یک تبدیل نیز روی فضای تصاویر اعمال شده و سپس دسته‌بند خطی یادگرفته شده

یادگیری w و v با استفاده از مسئله بهینه‌سازی زیر تعیین صورت می‌گیرد:

$$\min_{\mathcal{V}, \mathcal{W}, \xi, \epsilon} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{\lambda_1}{2} \sum_{\mathbf{v} \in \mathcal{V}} \|\mathbf{v}\|^2 + \lambda_2 \sum_{y,s} \epsilon_{ys} + \lambda_3 \sum_{i,y} \xi_{iy} \quad (36-2)$$

$$\text{s.t. } \forall i \in \{1, \dots, N\}, \forall y \in \mathcal{S}, \forall s \in \mathcal{S},$$

$$\sum_{i=1}^N \frac{\mathbb{I}_{\{y_i=y\}}}{N_y} [f(\mathbf{x}_i, y) - f(\mathbf{x}_i, s)] \geq \Delta(y, s) - \epsilon_{ys}, \quad (37-2)$$

$$f(\mathbf{x}_i, y_i) - f(\mathbf{x}_i, y) \geq \Delta(y_i, y) - \xi_{iy}, \quad (38-2)$$

$$\epsilon_{ys} \geq 0, \xi_{iy} \geq 0, \forall \mathbf{v} \in \mathcal{V}, \mathbf{v} \geq 0,$$

که در آن $\Delta(\cdot, \cdot)$ یک تابع هزینه‌ی خطای ساختارمند میان دسته‌ی پیش‌بینی شده و دسته‌ی صحیح را نشان می‌دهد $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ and $\lambda_3 \geq 0$ فرامترهای مربوط به منظم‌سازی هستند و $\xi = \{\xi_{iy}\}$ and $\epsilon = \{\epsilon_{ys}\}$ متغیرهای مربوطه به محدودیت‌های نرم در بهینه‌سازی‌اند. در این روش تابع هزینه‌ی خطای ساختارمند به صورت $\Delta(y, s) = 1 - \mathbf{c}_y^T \mathbf{c}_s$ تعریف شده است.

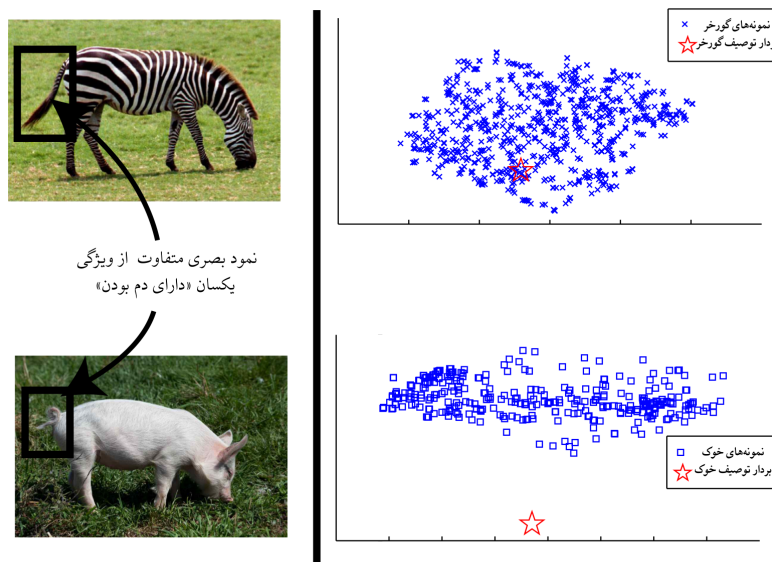
صورت‌بندی بالا یک صورت‌بندی دسته‌بندی با بیشترین حاشیه است با این تفاوت که علاوه بر محدودیت بیشترین حاشیه (رابطه (۲-۳۸)) یک محدودیت برای دسته‌بندی صحیح به صورت میانگین هم در رابطه (۲-۳۷) اضافه شده است. این محدودیت جدید می‌تواند باعث شود که داده‌ها به گونه‌ای نگاشته شود که نه تنها دسته‌بندی صحیح صورت گیرد بلکه یک توزیع با مرکز $\theta(c^y)$ ایجاد کنند و برای نگاشت‌یافته‌ی مراکز دسته‌ها نیز یک حاشیه در نظر گرفته شود. این حالت باعث ایجاد خوشه‌هایی جدا از هم می‌شود که مراکزشان توصیف‌هاست و در نتیجه برای مسئله یادگیری از صفر مناسب‌تر است.

نویسندگان این پژوهش روش خود را در [۳۸] با یادگیری توانان نگاشت توصیف‌ها و تصاویر توسعه داده‌اند. علاوه بر یادگیری توانان پارامترهای نگاشت‌ها، برای داده‌های تست، نمایش طوری به دست می‌آید که علاوه بر هم‌خوانی با پارامترهای بدست آمده برای نگاشت، از داده‌های دسته‌های دیده شده نیز دور باشند. این یک شرط شهودی برای بهتر شدن نگاشت است چرا که فرض بر این است که دسته‌های آموزش و آزمون اشتراکی ندارند و در نتیجه برای مثال نمایش تصاویر آزمون نباید در نزدیکی توصیف دسته‌های آموزش باشد.

۸-۲ روش‌های نیمه‌نظارتی

در این بخش به بررسی روش‌های نیمه‌نظارتی می‌پردازیم. این روش‌ها از نظر نوع نگاشت‌های مورد استفاده در یکی از دسته‌های قبلی قابل بیان بودند ولی با توجه به این که روش پیشنهادی ما نیز نیمه‌نظارتی است، برای پررنگ‌تر شدن نحوه‌های استفاده از داده‌های آزمون در جریان آموزش این دسته را به طور جداگانه مورد بررسی قرار می‌دهیم.

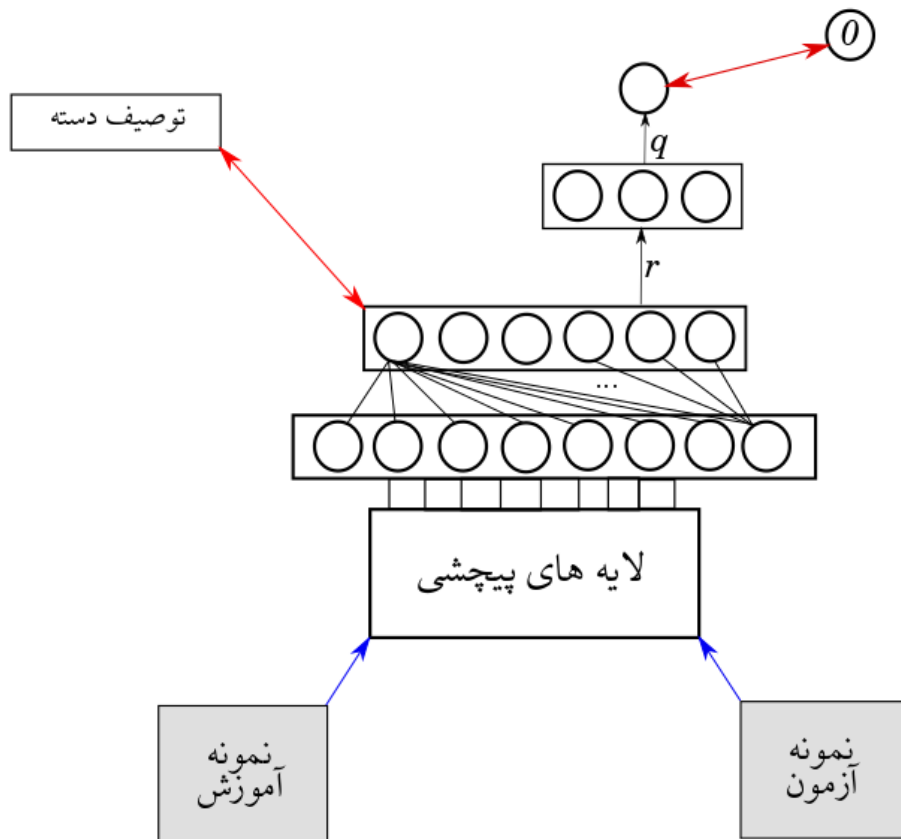
در [۳۹] برای نخستین بار مشکل جابجایی دامنه^{۳۷} معرفی شد. این مشکل که در شکل ۲-۳ قابل مشاهده است به تفاوت بودن خواص ویژگی‌ها برای دسته‌های مختلف اشاره می‌کند. برای مثال ویژگی راه‌راه بودن برای دو حیوان گورخر و ببر از نظر بصری خواص متفاوتی دارد و یادگیری یک دسته‌بند برای تشخیص راه‌راه بودن با استفاده از تصاویر گورخر در تشخیص وجود و یا عدم وجود این ویژگی در تصویر ببر ضعیف خواهد بود.



شکل ۲-۳: مشکل جابجایی دامنه بین دو دسته‌ی دیده شده (گورخر) و دیده نشده (خوک) نمایش داده شده است. ویژگی یکسان «دارای دم بودن» در این دو دسته دارای دو نمود بصری متفاوت است (سمت چپ) و نگاشت یادگرفته شده برای بردن این ویژگی به فضای مشترک برای دسته‌ی دیده نشده عمل‌کرد ضعیف‌تری نسبت به دسته‌ی دیده شده به نمایش می‌گذارد (سمت راست) [۳۹].

در [۳۹] برای حل این مشکل دو تکنیک به کار گرفته شده است. ابتدا یافتن نمایش مشترک برای سه دامنه‌ی تصاویر، بردار ویژگی و بردار نام دسته‌ها به صورت توامان با استفاده از CCA^{38} [۴۰] سپس برچسب‌گذاری داده‌های بدون برچسب

^{۳۷}Domain shift problem^{۳۸}Canonical Correlation Analysis



شکل ۳-۱: ساختار شبکه چند وظیفه‌ای پیشنهادی. فلش‌های آبی رنگ ورودی‌های شبکه را نشان می‌دهند و فلش‌های قرمز رنگ مقایسه خروجی شبکه با خروجی مورد انتظار را. خطوط سیاه رنگ اتصالات شبکه را نشان می‌دهند. زیر شبکه‌ی برگرفته شده از شبکه vgg و یک لایه‌ی با اتصالات چگال اضافه شده بین دو ورودی مشترک هستند. لایه‌های r و q مخصوص نمونه‌های آزمون هستند. خروجی لایه‌ی q همواره با مقدار صفر مقایسه می‌شود.

۳-۱ یک شبکه عصبی چندوظیفه‌ای

یادگیری نگاشت‌ها با استفاده از داده‌های دسته‌های دیده‌شده، همان‌طور که در بخش ۲-۸ اشاره شد، دچار مشکل جابجایی دامنه است و روی داده‌های دسته‌های دیده‌نشده به خوبی قابل تعمیم نیست. یک راه حل برای مقابله با این مشکل این است که در حین یادگیری نگاشت اجبار شود که حاصل نگاشت یک نمونه‌ی آزمون به نوعی نزدیک به نگاشت توصیف دسته‌های آزمون باشد. همان‌طور که در بخش ۲-۸ بیان شد، چنین راه‌حلی در [۴۲] استفاده شده است. معیار نزدیکی

نگاشت‌ها در آن روش ه یک امتیاز پیشین از شباهت هر نمونه‌ی آزمون با دسته‌های دیده نشده است که توسط یک روش دیگر استخراج شده می‌شود. ما در این بخش یک روش مبتنی بر شبکه‌های عصبی عمیق معرفی می‌کنیم که در آن نگاشتی غیرخطی و چندلایه از تصاویر به بردارهای ویژگی یادگرفته می‌شود. یادگیر نگاشت با معیار پیش‌بینی صحیح ویژگی برای نمونه‌های آموزش (که بردار ویژگی صحیح برای آن‌ها مشخص است) و هم‌چنین نزدیک بودن حاصل نگاشت هر نمونه‌ی آزمون به توصیف یکی از دسته‌های دیده نشده است. برای مدل کردن این نگاشت، از یک شبکه‌ی عصبی استفاده شده است. اگر نگاشت مدل شده با شبکه عصبی را با f نشان دهیم، تابع هزینه‌ی مورد استفاده برای آموزش شبکه به صورت زیر تعریف می‌شود:

$$\min_f \frac{1}{N_s} \sum_{n=1}^{N_s} \|f(\mathbf{x}_i) - \mathbf{c}_{y_i}\|_2^2 + \frac{\beta}{N_u} \sum_{i=N_s+1}^{N_s+N_u} \left(\min_{j=n_s, \dots, n_s+n_u} \|f(\mathbf{x}_i) - \mathbf{c}_j\|_2^2 \right), \quad (1-3)$$

که β یک فرایارامتر است. جمله‌ی اول، جمله‌ی مرسوم خطای میانگین مربعات است که در پیش‌بینی مقادیر پیوسته به کار گرفته می‌شود. جمله‌ی دوم برای رفع مشکل جابجایی دامنه طراحی شده است و تحمیل می‌کند که حاصل نگاشت یک نمونه‌ی آزمون حتماً نزدیک توصیف یکی از دسته‌های دیده‌نشده باشد، این دسته‌ی دیده نشده، دسته‌ای در نظر گرفته شده است توصیف آن با نگاشت کمترین فاصله را دارد. این انتخاب بر این فرض استوار است که حاصل نگاشت در اکثر موارد به توصیف صحیح نزدیکتر است، یا به عبارتی این که در اکثر مواقع استفاده از دسته‌بند نزدیکترین همسایه روی نگاشتی که تنها با جمله‌ی اول آموزش دیده، دقتی بیش از ۵۰٪ دارد. با این فرض، جمله‌ی دوم این تابع هزینه باعث می‌شود که حاصل روی نمونه‌های آزمون نیز بهتر شود و نگاشته‌شده‌ی هر تصویر فاصله کمتری با توصیف صحیح خود داشته باشد. تابع \min به کاربرده شده در جمله‌دوم معادله (۱-۳) در برخی نقاط مشتق‌پذیر نیست، اما با توجه به اینکه اندازه‌ی این نقاط صفر است تابع تقریباً همه‌جا مشتق‌پذیر است و آموزش شبکه با استفاده از پس‌انتشار^۱ مقدار گرادیان ممکن خواهد بود. به صورت دقیق‌تر، بهینه‌سازی رابطه (۱-۳) عملیات محاسبه‌ی مقدار کمینه را داخل شبکه تعبیه می‌کنیم (شکل ۱-۳)؛ به این صورت که لایه‌های جدید q و r برای نمونه‌های دیده نشده اضافه می‌شود که:

$$(q(\mathbf{x}))_j = \|f(\mathbf{v}) - \mathbf{c}_j\|_2^2, \quad (2-3)$$

$$r(\mathbf{z}) = \min_{j=1 \dots n_u} (\mathbf{z})_j. \quad (3-3)$$

^۱back propagation

در هنگام آموزش با پس انتشار، مشتق تابع هزینه l نسبت به هر ورودی مثل z در لایه r با ضابطه زیر محاسبه می شود:

$$\frac{\partial l}{\partial z} = \sum_j \mathbb{1}[(z)_j = \min(z)] \cdot \frac{\partial l}{(z)_j}. \quad (۴-۳)$$

پس از آموزش شبکه، در فاز آزمون لایه های q و r حذف شده و بردار توصیف برای تصاویر آزمون با استفاده از شبکه پیش بینی می شود، در نهایت دسته بندی با استفاده از دسته بند نزدیک ترین همسایه روی نمونه های آزمون انجام خواهد شد:

$$y_n^* = \underset{j}{\operatorname{arg\,min}} \|f(\mathbf{x}_n - \mathbf{c}_j)\|_2^2 \quad (۵-۳)$$

۳-۱-۱ معماری شبکه

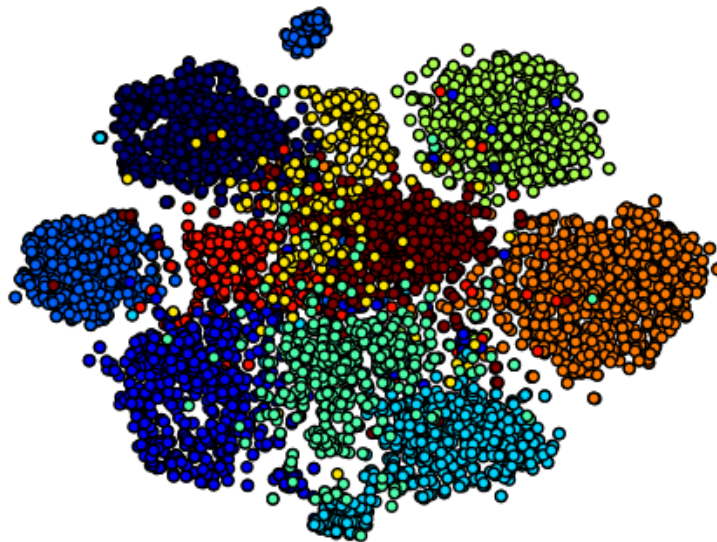
ما از قسمتی از شبکه ی ۱۹ لایه ی vgg [۱۴] که شامل ۱۶ لایه ی پیچشی ابتدا و لایه اول با اتصالات چگال به عنوان یک زیر شبکه در ورودی شبکه خود استفاده می کنیم. با این زیر شبکه تصاویر ورودی به بردارهای ۴۰۹۶-بعدی نگاشته می شوند. سپس یک لایه ی با اتصالات چگال قرار دارد که این حاصل را به بردارهای توصیف دسته ها می نگارد. سپس برای نمونه های آموزش خروجی این لایه با بردار توصیف صحیح مقایسه می شود. برای نمونه های آزمون خروجی این لایه به لایه های q و r متصل می شود و مقدار خروجی r با مقدار مطلوبش که صفر است مقایسه خواهد شد. توابع فعال سازی در همه ی لایه ها تابع ReLU است که

$$ReLU(\mathbf{x}) = \max(0, \mathbf{x}), \quad (۶-۳)$$

البته در لایه ی با اتصالات چگال آخر برای مجموعه دادگانی که مقادیر بردار توصیف دودویی هستند از تابع سیگموئید برای فعال سازی استفاده شده است تا مقادیر در بازه ی $[0, 1]$ نگاشته شوند.

۳-۲ تابع مطابقت مبتنی بر خوشه بندی

در اکثر روش های پیشین که در فصل ۲ مرور شد، تابع مطابقت میان تصاویر و توصیف ها برای اختصاص برجسب به داده های آزمون بر اساس فاصله کمینه یا ضرب داخلی بیشینه در یک فضای مشترک محاسبه می شد. استثنای این



شکل ۳-۲: نمایش دوبعدی بوسیله t -SNE برای ده دسته‌ی آزمون از مجموعه دادگان AWA با ده رنگ متفاوت نشان داده شده است. درستی فرض قابل خوشه‌بندی در تصویر مشخص است، یعنی ویژگی‌های استخراج شده با استفاده از شبکه‌های عمیق توانایی ایجاد تمایز بالا میان دسته‌ها را دارا هستند.

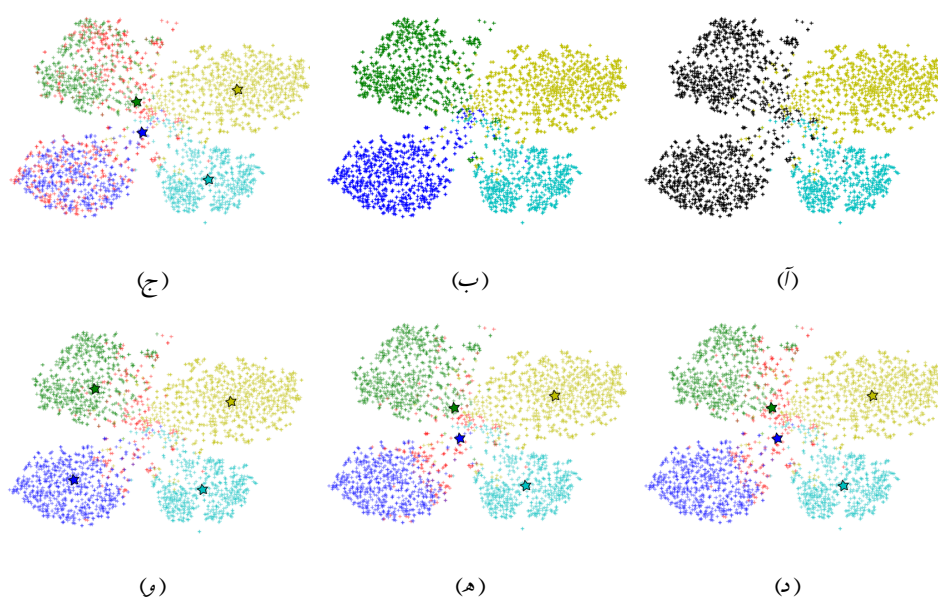
موضوع، استفاده از روش انتشار برچسب در [۳۹] و [۴۲] و همچنین پیش‌بینی مستقیم برچسب‌ها در [۴۱] و [۴۵] هستند.

در این بخش ما یک تابع مطابقت جدید بر اساس یک خوشه‌بندی روی داده‌های دسته‌های دیده نشده، تعریف می‌کنیم. اگر فضای نمایش تصاویر دارای این خاصیت باشد که دسته‌های مختلف به صورت خوشه‌های مجزا باشند، استفاده از خوشه‌بندی برای دسته‌بندی برای انتساب برچسب از نظر شهودی توجیه‌پذیر است. با توجه به نمایش غنی بوجود آمده برای تصاویر توسط شبکه‌های عمیق این فرض در بسیاری از موارد برقرار است. برای نمونه، نمایش t -SNE نمونه‌های آزمون مجموعه داده‌های AWA در تصویر ۳-۲ نشان داده شده است و برقراری فرض قابل خوشه‌بندی بودن در آن قابل مشاهده است. این ادعا با استفاده از آزمایش در بخش ۴-۴ اثبات خواهد شد. روش‌های پیشنهادی ما در این فصل بر اساس این ساختار و استفاده از وجود چنین خاصیتی در فضای تصاویر است. یک راه استفاده از چنین خاصیتی در فضای تصاویر، معرفی یک تابع مطابقت است که علاوه بر شباهت نگاشت‌یافته‌ی نمونه‌ها و توصیف‌ها به سایر نمونه‌های در همسایگی هر نمونه نیز وابسته باشد. بدین منظور ما یک تابع مطابقت جدید پیشنهاد می‌دهیم که در آن برچسب تعلق گرفته به هر

جدول ۴-۴: مقایسه دقت دسته‌بندی چنددسته‌ای روش پیشنهادی با سایر روش‌ها. نتایج بر اساس نوع ویژگی مورد استفاده برای تصاویر دسته‌بندی شده‌اند. جدول شامل دقت دسته‌بندی چنددسته‌ای به صورت (میانگین \pm انحراف معیار) است. نتایج سایر روش‌ها از مقالاتی که روش در آن‌ها ارائه شده نقل شده و آزمایش‌ها توسط ما تکرار نشده است. نتایج روش‌های پیشنهادی حاصل سه اجرا هستند.

ویژگی تصاویر	روش	Attributes with Animals	CUB-۲۰۱۱	aPascal-aYahoo	SUN
کم عمق	[۴۱] Li and Guo	$۳۸/۲ \pm ۲/۳$			$۱۸/۹ \pm ۲/۵$
	[۴۵] Li <i>et al.</i>	$۴۰/۰۵ \pm ۲/۲۵$		$۲۴/۷۱ \pm ۳/۱۹$	
	[۴۴] Jayaraman and Grauman	$۴۳/۰۱ \pm ۰/۰۷$		$۲۶/۰۲ \pm ۰/۰۵$	$۵۶/۱۸ \pm ۰/۲۷$
GoogleNet	[۲۸] Akata <i>et al.</i>	$۶۶/۷$	$۵۰/۱$		
	[۳۱] Xian <i>et al.</i>	$۷۱/۹$	$۴۵/۵$		
VGG-۱۹	[۴۲] Khodirov <i>et al.</i>	$۷۳/۲$	$۳۹/۵$	$۲۶/۵$	
	[۲۸] Akata <i>et al.</i>	$۶۱/۹$	$۵۰/۱$		
	[۳۷] Zhang and Saligrama	$۷۶/۳۳ \pm ۰/۵۳$	$۳۰/۴۱ \pm ۰/۲۰$	$۴۶/۲۳ \pm ۰/۵۳$	$۸۲/۵۰ \pm ۱/۳۲$
	[۳۸] Zhang and Saligrama	$۸۰/۴۶ \pm ۰/۵۳$	$۴۲/۱۱ \pm ۰/۵۵$	$۵۰/۳۵ \pm ۲/۹۷$	$۸۳/۸۳ \pm ۰/۲۹$
	پیشنهادی (ساده + k-means)	$۸۶/۳۴ \pm ۰/۱۳$	$۵۲/۴۸ \pm ۰/۶۰$	$۴۸/۰۳ \pm ۱/۵۶$	$۷۵/۷۵ \pm ۱/۰۶$
	پیشنهادی (ساده)	$۸۶/۳۸ \pm ۰/۵۶$	$۵۳/۱۰ \pm ۰/۴۳$	$۴۸/۰۰ \pm ۰/۶۹$	$۸۰/۶۶ \pm ۰/۷۶$
	پیشنهادی (توام، مقداردهی D)	$۸۳/۰۳$	$۵۷/۵۵$	$۴۲/۶۲$	$۷۲/۵۰$
	پیشنهادی (توام، مقداردهی R)	$۸۸/۶۴ \pm ۰/۰۴$	$۵۸/۸۰ \pm ۰/۶۴$	$۴۹/۷۷ \pm ۲/۰۲$	$۸۶/۱۶ \pm ۰/۵۷$

در بخش ۳-۲ برای دسته‌بندی بسیار موفق‌تر از دسته‌بند نزدیک‌ترین همسایه عمل می‌کند و اطلاعات غیر نظارتی موجود در نمونه‌های آزمون دقت دسته‌بندی را بهبود می‌دهد. هم‌چنین برتری روش خوشه‌بندی پیشنهادی در تصویر ۴-۱ قابل مشاهده است. در تصاویر ۴-۱ ج تا ۴-۱ ه که از نگاشت (۳-۹) برای تصویر کردن توصیف‌ها در فضای تصاویر استفاده شده است، مشکل جابجایی دامنه کاملاً قابل رویت است، یعنی برای دسته‌های دیده شده توصیف‌ها به صورت مناسبی در مرکز نمونه‌های آن دسته نگاشته شده‌اند حال آن که برای دسته‌های دیده نشده جابجایی وجود دارد و توصیف‌های آن‌ها از نمونه‌هاشان فاصله گرفته‌اند؛ اما در تصویر ۴-۱ و که از روش خوشه‌بندی و یادگیری نگاشت توام استفاده شده است این مشکل برطرف شده است و توصیف‌های دسته‌های دیده نشده نیز مانند دسته‌های دیده شده به مرکز نمونه‌های مربوط به خودشان نگاشته شده‌اند.



شکل ۴-۱: نمایش دوبعدی چهار دسته از مجموعه دادگان AWA با استفاده از نگاشت t -SNE، دو دسته‌ی دیده شده شامل بزگوزن (فیروزه‌ای) خرس گریزلی (زرد) و دو دسته‌ی دیده نشده شامپانزه (آبی) و پاندا (سبز). تصاویر با نماد بعلاوه و نگاشت توصیف دسته‌ها در فضای تصاویر با ستاره نشان داده شده است. در تصاویر b تا f نقطه‌های قرمز نمونه‌هایی که را نشان می‌دهد که دسته‌ای به جز چهار دسته‌ی موجود در شکل برای آن‌ها پیش‌بینی شده است. (آ) دسته‌های دیده شده با برچسب صحیح و دیده نشده با رنگ مشکی (ب) نمایش برچسب صحیح برای تمامی دسته‌ها (ج) توصیف‌ها با نگاشت (۳-۹) به فضای تصاویر برده شده‌اند و دسته‌بندی با دسته‌بند نزدیک‌ترین همسایه انجام شده است. (د) نگاشت مانند حالت قبل و دسته‌بندی با تابع مطابقت پیشنهادی به همراه خوشه‌بند k -means (ه) نگاشت مانند حالت قبل و دسته‌بندی با تابع مطابقت پیشنهادی به همراه خوشه‌بند نیمه‌نظارتی پیشنهاد شده (و) دسته‌بندی و نگاشت با استفاده از روش پیشنهادی برای یادگیری نگاشت و خوشه‌بندی توام