سجاد ایوبی
و
علیرضا سلطانی نشان

بهینه کردن هزینه حاشیه ای برای تشخیص چهره با یادگیری عمیق

مباحث ویژه در برنامه نویسی
استاد: برهلیا
۵ دی سال ۹۹

چکیده

تشخیص چهره[1] در چند سال اخیر به نتایج قابل ملاحظه ای دست یافته است که دلیل عمده آن پیشرفت سریع شبکه های عصبی عمیق[2] (DNN) است. تابع هزینه در یک شبکه عصبی عمیق است که منجر به ایجاد عملکرد متفاوتی می شود. در چند سال اخیر بعضی از توابع هزینه پیشنهاد داده شده است که با اینکه ادعا می کنند،اما در عمل نمی توانند مسئله بهینه کردن هزینه حاشیه ای[3] را که در مجموعه داده های پیچیده وجود دارد حل کنند. در این مقاله حل مسئله تمایل به بهینه کردن هزینه حاشیه ای را با در نظر گرفتن یک حاشیه حداقلی برای تمامی کلاس ها پیشنهاد می کنیم. ما تابع هزینه جدیدی به اسم حداقل هزینه حاشیه ای[4] (MML) ارائه می کنیم که هدف آن گسترش بازه ای داده هایی است که به نمونه های مرکزی کلاس خود بیش از اندازه نزدیک می شوند تا قابلیت دسته بندی کننده ویژگی های عمیق را تقویت کنند. تابع MML همراه با توابع Softmax Loss و Centre Loss بر فرآیند آموزش شبکه عصبی نظارت می کنند تا حاشیه های تمامی کلاس ها را صرف نظر از توزیع کلاس ها دسته بندی کنند. ما تابع MML را در معماری Inception-ResNet-v1 پیاده سازی می کنیم و آزمایش ها را به طور کامل بر روی هفت مجموعه داده تشخیص چهره انجام تست می کنیم که عبارت اند از MegaFace، FaceScrub، LFW، SLLFW،

1 Face recognition
2 deep neural networks
3 margin bias
4 Minimum Margin Loss

YTF، IJB-B و IJB-C. نتایج تجربی ما نشان می دهد که تابع هزینه MML که ما پیشنهاد کردیم باعث می شود اثر منفی این که داده ها به حاشیه ای کلاس های خود تمایل داشته باشند به شدت کاهش یابد.

کلید واژگان :یادگیری عمیق، شبکه های عصبی کانولوشنی[5] (CNN)، تشخیص چهره، حداقل هزینه حاشیه ای (MML)

1. مقدمه

در یک دهه گذشته، روش های مبتنی بر شبکه عصبی عمیق (DNN) به پیشرفت بزرگی در حوزه بینایی کامپیوتر شامل تشخیص چهره [1]، باز شناسایی افراد [2]، تشخیص اشیا و تشخیص ارقام دست نویس یافته است. پیشرفت در زمینه تشخیص چهره به خاطر دو عامل مهم یعنی مجموعه داده های حجیم تر و پیچیده تر چهره و توابع هزینه بهتر به دست آمده است.

کمیت و کیفیت مجموعه دادگان چهره برای آموزش شبکه تاثیر مهمی بر بهبود عملکرد یک سیستم DNN در تشخیص چهره دارند. این مجموعه دادگان به شکل عمومی در دسترس قرار دارند که برخی از آن ها ،[6] VGG Face2 ،[5] MS-Celeb-1M [7] MegaFace و [8] CASIA WebFace هستند. همانطور که در جدول 1 مشاهده می کنید، مجموعه داده CASIA WebFace شامل تصاویر 0.5 M است؛ مجموعه داده VGG Face2 شامل تصاویر 3M چهره اما تنها از 9K انسان منحصر به فرد است. مجموعه داده MS-Celeb-1M و MegaFace شامل چهره ها و افراد های بیشتری

[5] Convolutional neural networks

است، به همین دلیل باید دارای توان بیشتری برای آموزش دادن یک مدل DNN باشد. از این گذشته، هر دو مجموعه داده MS-Celeb-1M و MegaFace دارای مشکل توزیع long-tailed هستند [29] که یعنی افراد خیلی کمی دارای تعداد زیادی از تصاویر چهره هستند و تعداد زیادی از افراد تصاویر چهره بسیار محدودی دارند. زمانی که از مجموعه دادگانی با توزیع long-tailed استفاده می کنیم، مدل آموزش دیده تمایلی بیش از حد به کلاس های دارای نمونه زیاد دارد که این باعث ضعیف شدن تعمیم شبکه عصبی در بخش long-tailed می شود [9]. به خصوص، کلاس هایی که دارای نمونه های فراوانی هستند تمایل دارند که فاصله نسبتا زیادی با نمونه های مرکزی کلاس خود داشته باشند؛ همچنین، کلاس هایی که دارای نمونه های کمتری هستند علاقه دارند تا فاصله نسبتاً کمی با نمونه های مرکزی کلاس خود داشته باشند چون که آن ها منطقه کوچکی را در فضا اشغال می کنند و در نتیجه به آسانی فشرده می شوند. این مسئله ای تمایل به فشرده شدن به دلیل توزیع long-tailed برخی کلاس هاست که منجر به ضعیف شدن عملکرد در تشخیص چهره می شود [9].

جدول 1: داده های مربوط به مجموعه داده های عمومی موجود برای چهره.

	MS-Celeb-1M	VGGFace2	MegaFace	CASIA
تعداد هویت ها	100K	9K	672K	11K
تعداد تصاویر	10M	3M	5M	0.5M
متوسط به ازای هر فرد	105	323	7	47

علاوه بر مجموعه آموزش و توزیع آماری کلاس ها، عامل مهم دیگری بر عملکرد تابع هزینه تاثیرگذار است که شبکه برای بهینه سازی وزن آن در طی فرآیند آموزش بهینه می کند. بهترین عملکرد توابع هزینه فعلی به دو دسته تقسیم می شود: توابع هزینه

مبتنی بر فاصله اقلیدسی[6] و توابع هزینه مبتنی بر فاصله کسینوسی[7]. بیشتر آن ها از تابع Softmax Loss با افزودن یک جریمه یا تغییر مستقیم softmax گرفته شده اند. توابع هزینه مبتنی بر فاصله اقلیدسی شامل توابع Triplet ،[10] Contrastive Loss [9] Rane Loss ،[12] Centre Loss ،[11] Loss و [13] Marginal Loss است. هدف این توابع بهبود توانایی تمایز ویژگی ها با حداکثر کردن فاصله بین دسته ها یا حداقل کردن فاصله درون دسته است. تابع Contrastive Loss نیازمند این است که شبکه دو نوع زوج نمونه را به عنوان ورودی بگیرد: زوج نمونه های مثبت (دو چهره از یک دسته) و زوج نمونه های منفی (دو تصویر چهره از دسته های مختلف). تابع Contrastive Loss فاصله اقلیدسی زوج های مثبت را به حداقل می رساند و زوج های منفی که دارای فاصله کمتر از یک آستانه هستند را جریمه می کند. تابع Triplet Loss از یک سه تایی به عنوان ورودی استفاده می کند که شامل یک نمونه مثبت، یک نمونه منفی و یک لنگر[8] است. لنگر نیز یک نمونه مثبت است که در ابتدا به برخی نمونه های منفی نسبت به برخی نمونه های مثبت نزدیک تر است. در طی مرحله آموزش، زوج های لنگر- مثبت با هم گرفته می شوند در حالی که زوج های لنگر- منفی تا اندازه ممکن از یکدیگر دور می شوند. با این حال، انتخاب زوج های نمونه و سه تایی ها برای توابع Contrastive Loss و Triplet Loss پر زحمت و وقت گیر هستند. توابع Marginal Loss و Rane Loss جریمه دیگری را برای پیاده سازی نظارت مشترک توسط تابع Softmax Loss اضافه می کنند. به ویژه، Center Loss جریمه ای

6 Euclidean distance
7 Cosine distance
8 anchor

را با محاسبه و محدود کردن فواصل بین نمونه های درون دسته و مرکز دسته مربوطه می افزاید. تابع Marginal Loss تمامی زوج های نمونه در یک دسته را در نظر می گیرد و زوج های نمونه مربوط به دسته های مختلف را مجبور می کند تا حاشیه بزرگتری نسبت به آستانه θ در اختیار داشته باشند، در حالی که نمونه های مربوط به یک دسته را مجبور می کند تا حاشیه کوچکتری نسبت به آستانه θ داشته باشند. با این حال، دو نمونه دور در یک دسته را مجبور می کند تا فاصله کمتری نسبت به دو نمونه نزدیک از دسته های مختلف داشته باشند که سبب دشوار شدن همگرایی روند آموزش می شود. تابع Range Loss فواصل نمونه ها در هر دسته را محاسبه می کند و زوج دو نمونه ای را انتخاب می کند که دارای بیشترین فاصله به عنوان محدودیت درون دسته ای هستند؛ به طور همزمان، تابع Range Loss فاصله هر زوج مرکز دسته (یعنی زوج مرکز) را محاسبه می کند و زوج مرکز که دارای کمترین فاصله است مجبور می کند تا حاشیه بیشتری نسبت به آستانه طراحی شده داشته باشد. با این حال، تنها در نظر گرفتن مرکز هر زوج کافی نیست چرا که زوج های مرکز دارای حاشیه های کوچکتری نسبت به آستانه طراحی شده هستند و در نتیجه همگرایی کامل روند آموزش به دلیل سرعت پایین آموزش دشوار است.

توابع هزینه مبتنی بر فاصله کسینوسی شامل L-Softmax، L_2-Softmax Loss [14] AM-Softmax Loss [17]، A-Softmax-Loss [16]، Loss [15] و ArcFace [18] است. براساس هزینه Softmax، تابع L_2-softmax Loss نرم[9] L2 مربوط به توصیف کننده ویژگی را به یک مقدار ثابت محدود می کند. تابع L2-Softmax Loss تفسیر

[9] Norm

هندسی بهتری را ارائه می کند و توجه مشابهی به چهرها با کیفیت خوب و بد ارائه می کند. تابع L-Softmax خروجی لایه softmax را از تبدیل W.f به W|.|f|.cos θ| مجدداً فرموله می کند تا فاصله اقلیدسی را به فاصله کسینوسی تبدیل و همچنین ثابت های زاویه ای را برای افزایش حاشیه های زاویه ای بین هویت های مختلف اضافه می کند. براساس تابع L-Softmax Loss، تابع A-Softmax نرمالیزه کردن وزن را اضافه می کند، بنابراین W.f به f|cos θ| فرموله می شود که هدف آموزش را ساده یم کند. با این حال، پس از استفاده از ثابت های زاویه ای، همگرایی هر دو تابع L-Softmax و A-Softmax Loss دشوار می شود. بنابراین استراتژی بهینه سازی بازپخت[10] توسط این دو روش پذیرفته می شود تا به همگرایی الگوریتم کمک کند. به منظور بهینه کردن همگرایی تابع A-Softmax، Wang و همکاران [17] تابع AM-Softmax را پیشنهاد می دهند که ثابت های زاویه ای را با ثابت های زاویه ای جمع شوند به نام تبدیل cos(mθ) به cosθ - m جایگزین می کند. علاوه بر این، تابع AM-Softmax نیز نرمالیزه کردن ویژگی را به کار می گیرد و ضریب مقیاس دهی عمومی به نام s=30 را معرفی می کند که W|.|f|=s| را ایجاد می کند. بنابراین هدف آموزش یعنی W|.|f|. cosθ| مجدداً به s. cosθ ساده می شود. همچنین تابع ArcFace نیز ثابت های زاویه ای را به کار می گیرد اما cos(mθ) را به cos(m+θ) تبدیل می کند که دارای تفسیر هندسی بهتری است. هر دو تابع AM-Softmax و ArcFace نرمالیزه سازی وزن و نرمالیزه سازی ویژگی را می پذیرند به طوری که تمامی ویژگی ها را برای قرارگیری بر روی یک فراصفحه می پذیرند. با این حال، آیا تمامی ویژگی ها را مجبور می کند تا بر روی یک

[10] annealing optimization strategy

فراصفحه قرار گیرند، به جای این که بر روی یک صفحه بزرگتر قرار گیرند؟ چرا و چگونه نرمالیزه سازی وزن و نرمالیزه سازی ویژگی از روند آموزش سود می برند؟ پاسخ آشکار به این پرسش ها دشوار است و برخی مستندات نشان می دهند که نرمالیزه سازی ویژگی "نرم" می تواند منجر به نتایج بهتر شود [10].

توابع هزینه فعلی احتمال تمایل به حاشیه را در نظر نمی گیرند. برای اصلاح این تمایل به حاشیه، تعیین یک حاشیه حداقلی برای تمامی زوج دسته ها را پیشنهاد می دهیم و سپس یک تابع هزینه مبتنی بر حداقل حاشیه را طراحی می کنیم. با الهام از توابع Softmax Loss، Center Loss و Marginal Loss، تابع هزینه جدیدیی به نام حداقل هزینه حاشیه ای[11] (MML) پیشنهاد می دهیم که هدف آن اجبار تمامی زوج های مرکز دسته به داشتن فاصله بیشتر نسبت به حداقل حاشیه تعیین شده است. متفاوت از تابع Rang Loss، روش LLM تمامی زوج های مرکز دسته "بدون صلاحیت" را جریمه می کند، به جای این که تنها زوج مرکزی را جریمه کند که دارای کمترین فاصله است. روش MML از موقیعت های مرکز به طور مستمراستفاده می کند که توسط تابع Center Loss به روز می شوند و فرآیند آموزش را توسط نظارت مشترک توسط توابع Softmax Loss و Center Loss هدایت می کند. براساس اطلاعات موجود، هیچ تابع هزینهی وجود ندارد که تعیین حداقل حاشیه بین مراکز دسته را در نظر بگیر، د. با این حال، وجود چنین محدودیتی برای اصلاح تمایل به حاشیه ایجاد شده توسط عدم توزان دسته در داده آموزش ضروری است. برای اثبات اثربخشی روش پیشنهادی، آزمایش هایی بر روی هفت مجموعه داده عمومی انجام می شود: Labelled Faces in the

[11] Minimum Margin Loss

Wild (LFW) [20]، Similar-looking LFW (SLLFW) [21]، YouTube Faces (YTF) [22]، Megaface [7]، FaceScrub [23]، IJB-B [24] و IJB-C [25]. نتایج به دست آمده نشان می دهد که روش MML پیشنهادی به عملکرد بهتر نسبت به توابع Range Loss، Centre Loss، Softmax Loss و Marginal Loss دست یافته است که تقریباً هیچ افزایشی در هزینه ایجاد نشده است. علاوه بر این، روش پیشنهادی به عملکرد رقابتی در مقایسه با بهترین روش های موجود دست یافته است.

2. از تابع **Softmax Loss** تا تابع حداقل هزینه حاشیه ای

2-1 Sotmax Loss و Center Loss

تابع Softmax Loss پر استفاده ترین تابع هزینه است که در ذیل ارائه شده است:

$$\mathcal{L}_S = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{W_{y_i}^T f_i + b_{y_i}}}{\sum_{j=1}^{K} e^{W_j^T f_i + b_j}} \qquad (1)$$

که N اندازه مجموعه دسته و K شماره دسته مربوط به مجموعه دسته است. $f_i \in R^d$ نشان دهنده ویژگی i امین نمونه متعلق به دسته y_i ام است، $W_j \in R^d$ نشان دهنده ستون j امن ماتریس وزن W در لایه نهایی کاملاً متصل و b_j عبارت تمایل مربوط به دسته j ام است. براساس رابطه (1) مشاهده می شود که تابع Softmax Loss برای به حداقل رساندن تفاوت ها بین برچسب های پیش بینی شده و برچسب های صحیح طراحی شده است که به بیان دیگر به این معنا است که تابع Softmax Loss تنها ویژگی ها را از دسته های مختلف در مجموعه آموزش تفکیک می کند، به جای این که ویژگی های متمایز کننده را یاد بگیرد. چنین هدفی برای امور مجموعه بسته مانند بیشتر سناریوهای کاربردی تشخیص اشیا و تشخیص رفتار مناسب است. اما سناریوهای

کاربردی تشخیص چهره در بیشتر موارد از نوع مجموعه باز هستند، بنابراین توانایی تمایز ویژگی ها تاثیر قابل توجهی بر عملکرد سیستم تشخیص چهره دارد. برای ارتقاء توانایی تمایز ویژگی ها، Wen و همکاران [12] تابع Center Loss را برای به حداقل رساندن فاصله درون دسته ای پیشنهاد داده اند که در ذیل ارائه شده است:

$$\mathcal{L}_C = \frac{1}{2}\sum_{i=1}^{N} ||f_i - c_{y_i}||_2^2 \qquad (2)$$

که cy_i نشان دهنده مرکز دسته مربوط به دسته y_i ام است. تابع Center Loss تمامی فواصل بین مراکز دسته و در درون نمونه های دسته را محاسبه می کند و در ترکیب با تابع Softmax Loss مورد استفاده قرار می گیرد:

$$\mathcal{L} = \mathcal{L}_S + \lambda\mathcal{L}_C \qquad (3)$$

$$= -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{W_{y_i}^T f_i + b_{y_i}}}{\sum_{j=1}^{K} e^{W_j^T f_i + b_j}} + \frac{\lambda}{2}\sum_{i=1}^{N} ||f_i - c_{y_i}||_2^2 \qquad (4)$$

که λ فراپارامتر برای توازن دو تابع هزینه است.

2-2 هزینه حاشیه ای و هزینه محدوده

پس از ترکیب تابع Softmax Loss با تابع Centre Loss، فشردگی درون دسته به طور قابل توجهی ارتقاء می یابد. اما استفاده تنها از Softmax Loss به عنوان یک شرط بین دسته ای کافی نیست چرا که تفکیک پذیری ویژگی ها را ترغیب می کند. بنابراین Deng و همکاران [13] تابع Marginal Loss را پیشنهاد داده اند که نظارت مشترک را توسط Softmax Loss در نظر می گیرد.

$$\mathcal{L} = \mathcal{L}_S + \lambda \mathcal{L}_{Mar} \quad (5)$$

$$\mathcal{L}_{mar} = \frac{1}{N^2 - N} \sum_{i,j,i \neq j}^{N} \left(\xi - y_{ij} \left(\theta - \left\| \frac{f_i}{||f_i||} - \frac{f_j}{||f_j||} \right\|_2^2 \right) \right)_+ \quad (6)$$

که f_i و f_j به ترتیب ویژگی های در نمونه های i و j ام در یک دسته است؛ $y_{ij} \in \{ \pm 1 \}$ نشان می دهد که f_i و f_j به یک دسته تعلق دارند، $(u)_+$ به صورت $\max(u,0)$ تعریف می شود، θ آستانه برای تفکیک زوج های مثبت و منفی است و ξ حاشیه خطا در کنار فراصفحه دسته بندی است.

تابع Marginal Loss تمامی ترکیب های ممکن زوج های نمونه در یک مجموعه دسته را در نظر می گیرد و آستانه θ را برای محدود کردن این زوج نمونه ها شامل زوج های مثبت و منفی مشخص می کنند. تابع Margin Loss فواصل زوج های مثبت را مجبور می کند تا به آستانه θ نزدیک شوند و در عین حال فواصل زوج های منفی را ملزم می کند تا از آستانه θ دورتر شوند. البته به‌کارگیری همان آستانه θ برای محدود کردن زوج های مثبت و منفی به طور همزمان مناسب نیست چرا که در اغلب موارد دو نمونه دور در یک دسته دارای فاصله بیشتری نسبت به دو نمونه نزدیک از دو دسته متفاوت اما نزدیک هستند. تغییر اجباری این وضعیت سبب می شود که همگرایی روند آموزش دشوار شود.

مشابه با روش های اشاره شده در بالا، تابع Range Loss پیشنهادی توسط Zhang و همکاران [9] نیز با softmax Loss به عنوان سیگنال های نظارتی رفتار می کند:

$$\mathcal{L} = \mathcal{L}_S + \lambda \mathcal{L}_R \quad (7)$$

متفاوت از تابع Marginal Loss، تابع Range Loss شامل دو هزینهف مستقل به نام $\mathcal{L}_{R_{intra}}$ و $\mathcal{L}_{R_{inter}}$ برای محاسبه به ترتیب هزینه درون دسته و بین دسته ای است.

$$\mathcal{L}_R = \alpha \mathcal{L}_{R_{intra}} + \beta \mathcal{L}_{R_{inter}} \quad (8)$$

که α و β دو وزن برای تنظیم تاثیر $\mathcal{L}_{R_{intra}}$ و $\mathcal{L}_{R_{inter}}$ هستند. از نظر ریاضی، $\mathcal{L}_{R_{intra}}$ و $\mathcal{L}_{R_{inter}}$ به صورت ذیل تعریف می شوند:

$$\mathcal{L}_{R_{intra}} = \sum_{i \subseteq K} \mathcal{L}^i_{R_{intra}} = \sum_{i \subseteq I} \frac{n}{\sum_{j=1}^{n} \frac{1}{D_{ij}}} \quad (9)$$

$$\mathcal{L}_{R_{inter}} = max(M - D_{Centre}, 0) \quad (10)$$

$$= max(M - ||\bar{x}_{\mathcal{Q}} - \bar{x}_{\mathcal{R}}||_2^2, 0) \quad (11)$$

که K شماره دسته در مجموعه دسته فعلی، D_{ij} بزرگترین فاصله j ام از زوج های نمونه در دسته i، D_{Centre} فاصله مرکزی دو دسته نزدیک در مجموعه دسته فعلی، $\bar{x}_{\mathcal{Q}}$ و $\bar{x}_{\mathcal{R}}$ نشان دهنده مراکز دسته مربوط به دسته $x_{\mathcal{Q}}$ و $x_{\mathcal{R}}$ است که دارای کمترین فاصله مرکزی هستند و M آستانه حاشیه ای است. $\mathcal{L}_{R_{intra}}$ تمامی زوج های نمونه در یک دسته را اندازه گیری می کند و n دسته نمونه را شناسایی می کند که دارای فاصله بیشتر برای ایجاد هزینه به منظور کنترل فشردگی درون دسته ای هستند. مطابق با توضیح ارائه شده در[13]، آزمایش ها نشان می دهند که n=2 بهترین انتخاب است. هدف $\mathcal{L}_{R_{inter}}$ ملزم کردن زوج مرکز دسته که دارای کمترین فاصله است، به داشتن حاشیه بیشتر حداکثر تا آستانه طراحی شده است. اما زوج های مرکزی بیشتری وجود دارد که ممکن است فاصله کمتری نسبت به آستانه طراحی شده داشته باشند. البته در نظر گرفتن تنها یک زوج مرکز کافی نیست چرا که منجر به طولانی شدن روند آموزش برای تکمیل همگرایی می شود که دلیل آن سرعت پایین یادگیری است.

3-2 حداقل هزینه حاشیه ای پیشنهادی

با الهام از تابع Center Loss، Softmax Loss و Marginal Loss، حداقل هزینه حاشیه ای (MML) را در این مقاله پیشنهاد می دهیم. تابع MML همراه با Softmax Loss و Center Loss مورد استفاده قرار می گیرد که تابع Center Loss برای ارتقاء فشردگی مورد استفاده قرار می گرد و توابع Softmax و MML برای بهبود قابلیت تفکیک بین دسته به کار گرفته می شوند. به ویژه، Softmax مسئول تضمین صحت دسته بندی است، در حالی که هدف MML بهینه سازی حاشیه های بین دسته ای است. کل هزینه در ذیل نشان داده شده است:

$$\mathcal{L} = \mathcal{L}_S + \alpha \mathcal{L}_C + \beta \mathcal{L}_M \qquad (12)$$

که α و β فرا پارامترها برای تنظیم فشردگی Center Loss و MML است.

تابع MML یک آستانه به نام حداقل حاشیه (Minimum Margin) را مشخص می کند. با استفاده مجدد از موقعیت مرکز دسته که توسط Center Loss به روز می شوند، تابع MML تمامی زوج های مرکز دسته را براساس Minimum Margin مشخص پالایش می کند. برای زوج هایی که دارای فاصله کمتر از آستانه هستند، جریمه های مرتبط به درون مقدار هزینه افزوده می شوند. جزئیات مربوط به تابع MML به صورت ذیل فرموله می شود:

$$\mathcal{L}_M = \sum_{i,j=1}^{K} max(||c_i - c_j||_2^2 - \mathcal{M}, 0) \qquad (13)$$

که K شماره دسته یک مجموعه دسته و c_i و cj به ترتیب نشان دهنده i امین و j ایمن دسته ها هستند و M نشان دهنده حداقل حاشیه طراحی شده است. در هر مجموعه دسته آموزش، مراکز دسته توسط Center Loss توسط دو رابطه ذیل به روز می شوند:

$$c_j^{t+1} = c_j^t - \gamma \Delta c_j^t \quad (14)$$

$$\Delta c_j^t = \frac{\sum_{i=1}^{m} \delta(y_i = j)(c_j - f_i)}{1 + \sum_{i=1}^{m} \delta(y_i = j)} \quad (15)$$

که γ نرخ یادگیری مراکز دسته، t تعداد تکرارها و $\delta(\text{condition})$ یک تابع شرطی است. اگر شرط برقرار باشد، آنگاه $\delta(\text{condition}) = 1$ و در غیر اینصورت $\delta(\text{condition}) = 0$ است. لطفاً توجه داشته باشید که Range Loss مرکز یک دسته با متوسط گیری از نمونه های این دسته در یک مجموعه دسته محاسبه می شود. با این حال، اندازه یک مجموعه دسته محدود است و تعداد نمونه یک دسته خاص بسیار محدود است. بنابراین، مراکز دسته تولید شده به این صورت در مقایسه با مراکز دسته واقعی دقیق نیست. در مقایسه با تابع Range Loss، مراکز دسته یادگرفته شده مربوط به MML به مراکز دسته واقعی نزدیکتر هستند.

الگوریتم 1 مراحل اصلی یادگیری در شبکه های CNN با توجه به $\mathcal{L}_S + \mathcal{L}_C + \mathcal{L}_M$ پیشنهادی را نشان می دهد.

الگوریتم 1: الگوریتم یادگیری در شبکه های CNN همراه با $\mathcal{L}_S + \mathcal{L}_C + \mathcal{L}_M$ پیشنهادی.

ورودی: نمونه های آموزش {f_i}، پارامترهای مقداردهی شده اولیه θ_C در لایه های کانولوشن، پارامترهای W در لایه کاملاً متصل و n مرکز دسته مقدار دهی شده {$c_j \mid j = 1, 2, \ldots, n$}، نرخ یادگیری μ^t، فرا پارامترهای α و β، نرخ بادگیری مراکز دسته γ و تعداد تکرار .

خروجی: پارامترهای θ_C

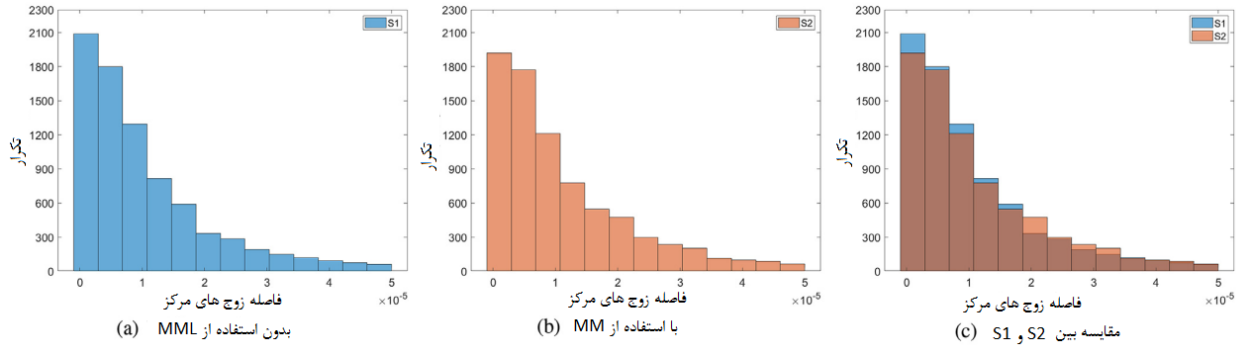
1: **while** not converge **do**
2: محاسبه حداکثر اتلاف $\mathcal{L}^t = \mathcal{L}_S^t + \alpha\mathcal{L}_C^t + \beta\mathcal{L}_M^t$.
3: محاسبه خطای انتشار $\frac{\partial \mathcal{L}^t}{\partial f_i^t}$ برای هر نمونه i
by $\frac{\partial \mathcal{L}^t}{\partial f_i^t} = \frac{\partial \mathcal{L}_S^t}{\partial f_i^t} + \alpha\frac{\partial \mathcal{L}_C^t}{\partial f_i^t} + \beta\frac{\partial \mathcal{L}_M^t}{\partial f_i^t}$.
4: به روز رسانی توسط $W^{t+1} = W^t - \mu^t \frac{\partial \mathcal{L}^t}{\partial W^t} = W^t - \mu^t \frac{\partial \mathcal{L}_S^t}{\partial W^t}$.
5: به روز رسانی c_j برای هر مرکز j by $c_j^{t+1} = c_j^t - \gamma\Delta c_j^t$.
6: به روز رسانی θ_C by $\theta_C^{t+1} = \theta_C^t - \mu^t \sum_i^N \frac{\partial \mathcal{L}^t}{\partial f_i^t}\frac{\partial f_i^t}{\partial \theta_C^t}$.
7: $t \leftarrow t + 1$.
8: **end while**

4-2 بحث

1-4-2 آیا **MML** به طور صحیح می تواند فاصله های مربوط به نزدیکترین زوج مرکز دسته را که کوچکتر از حداقل حاشیه مشخص شده هستند افزایش دهد؟

برای تایید این نکته ما از مدل های عمیق آموزش دیده توسط روش 1 (Softmax Loss + Center Loss) و روش 2 (Softmax Loss + Center Loss + MML) برای استخراج ویژگی ها از تمامی تصاویر از نسخه اصلاح شده مجموعه داده VGGFace2 استفاده می کنیم [6]. جزئیات مربوط به این مجموعه داده پامسازی شده و فرآیند آموزش این دو مدل در بخش 3-1 ارائه شده است. تفاوت بین روش 1 و روش 2 این است که روش 2 تابع MML را به عنوان بخشی از سیگنال نظارت در نظر می گیرد اما روش 2 این طور نیست. با استفاده از ویژگی های استخراج شده می توانیم موقعیت مرکز برای هر دسته را محاسبه کنیم و سپس فاصله بین هر مرکز دسته و نزدیک ترین مرکز دسته همسایه را محاسبه کنیم. توزیع فواصل هر یک از این مراکز دسته در شکل

1 نشان داده شده است. شکل های a–1 و b–1 توزیع فاصله به ترتیب در روش 1 و روش 2 را نشان می دهند. شکل c–1 مقایسه ای بین روش 1 و روش 2 انجام می دهد که براساس آن می توانیم مشاهده کنیم که روش 2 دارای مقادیر کوچکتری در پنج بخش اول است در حالی که مقادیر بزرگتری را در باقیمانده بخش ها در اختیار دارد. این نشان می دهد که تابع MML فاصله برخی زوج های مرکز همسایه را بیشتر می کند و در نتیجه تعداد زوج مرکز هایی که دارای حاشیه بزرگتری هستند را افزایش می دهد.



شکل 1: برای هر دسته در مجموعه داده VGGFace2، نزدیک ترین دسته همسایه آن با مقایسه جایگاه مراکز مختلف دسته یافت می شود. شکل های a، b و c توزیع فواصل بین هر مرکز دسته و نزدیک ترین مرکز دسته مربوط به آن را نشان می دهد. شکل (a) توزیع در مورد استفاده از ویژگی های استخراج شده از روش 1 (بدون استفاده از MML) را نشان می دهد. شکل (b) توزیع در مورد استفاده از ویژگی های تولید شده توسط روش 2 (با استفاده از MML) را نشان می دهد. شکل (c) نتایج مقایسه بین (a) و (b) را نشان می دهد که S1 و S2 به ترتیب روش های 1 و 2 را نشان می دهند.

2-4-2 آیا **MML** می تواند عملکرد مدل درباره تشخیص چهره را بهبود دهد؟

برای پاسخ به این پرسش آزمایش های گسترده ای را براساس مجموعه داده های مختلفی مطابق با بخش 3 انجام می دهیم. این انواع آزمایش شامل تایید چهره، شناسایی چهره، تشخیص مبتنی بر تصویر و تشخیص مبتنی بر ویدیو است. نتایج حاصل نشان می دهد که مدل پیشنهادی می تواند روش های پایه را همانند بهترین روش های ممکن عمل کند.

3. آزمایش ها

در این بخش جزئیات پیاده سازی آزمایش ها را توصیف می کنیم، تاثیر پارامترهای β و M را مورد بررسی قرار می دهیم و عملکرد روش پیشنهادی را ارزیابی می کنیم. این ارزیابی ها در مجموعه داده های ،[20] LFW ،[23] FaceScrub ،[7] MegaFace [24] IJB-B ،[22] YTF ،[21] SLLFW و [25] IJB-C همراه با شناسایی چهره و تایید چهره انجام می شوند. شناسایی چهره و تایید چهره دو کار اصلی در تشخیص چهره است. هدف تایید چهره تایید این است که آیا دو چهره مربوط به یک فرد است یا خیر که یک دسته بندی باینری محسوب می شود. هدف شناسایی چهره تشخیص ID یک چهره و پاسخ به ID دقیق است که یک مساله چند دسته بندی است.

1-3 جزئیات آزمایش

داده آموزش: در تمامی آزمایش ها از مجموعه داده VGGFace2 [6] به عنوان داده آموزش استفاده می کنیم. برای اطمینان از دقت و اطمینان پذیری نتایج تجربی، تمامی تصاویر چهره هایی که ممکن بود با مجموعه داده ها هم پوشانی داشته باشند را حذف کردیم. از آنجایی که نویز برچسب در این مجموعه داده بسیار پایین است، هیچ گونه

پاکس سازی داده اعمال نشده است. مجموعه داده نهایی حاوی 3.05M تصویر چهره از 8K هویت است.

پیش پردازش داده ها: مجموعه داده [26] MTCNN به تمامی تصاویر چهره برای موقعیت یابی، تراز صورت و تشخیص چهره اعمال می شود. اگر تشخیص چهره بر روی یک تصویر آموزش با شکست مواجه شود، آنگاه به سادگی از آن صرف نظر می کنیم؛ اگر بر روی یک تصویر تست ناموفق باشد، آنگاه موقعیت های فراهم شده مورد استفاده قرار می گیرند. تمامی تصاویر آموزش و تست به تصاویر 160*160 RGB تنظیم شده اند. به منظور تقویت داده ها، چرخش افقی تصادفی را بر روی تصاویر آموزش به کار می گیریم. برای بهبود دقت تشخیص، ویژگی های مربوط به تصویر تست را الحاق می کنیم. توجه داشته باشید که هیچ گونه پاکسازی داده ها بر روی مجموعه های تست موجود در آزمایش ها از جمله مجموعه داده Megaface انجام نشده است.

تنظیمات شبکه: براساس مجموعه داده [27] Inception-ResNet-v1، پنج مدل را توسط [28] Tensorflow مطابق با پنج روش نظارت پیاده سازی و آموزش داده ایم: Softmax Loss، Softmax Loss + Center Loss، Softmax Loss + Marginal Loss, Softmax Loss + Range Loss و Softmax Loss + Centre Loss + MML. برای راحتی کار، از "Range Loss"، "Marginal Loss"، "Centre Loss"، "Softmax Loss" و "MML" برای نمایش این پنج روش به ترتیب در نتایج آزمایش استفاده می کنیم. ما این پنج مدل را براساس GPU (GTX 1080 Ti) آموزش می دهیم و 90 را به عنوان اندازه مجموعه داده، 512 را به عنوان اندازه گنجاندن، 5e-4 را به عنوان افت وزن و

0.4 را به عنوان احتمال لایه کاملاً متصل تعیین می کنیم. تعداد کل تکرارها برابر 275K است که حدود 30 ساعت به طول می انجامد. نرخ یادگیری به صورت 0.05 آغاز می شود و هر 100K بر 10 تقسیم می شود. تمامی روش ها از همان تنظیمات پارامتری استفاده می کنند، به استثنای Softmax Loss + Centre Loss + MML که مدل آموزش دیده Softmax Loss + Centre Loss را به عنوان مدل پیش آموزش دیده بارگذاری می کند پیش از آن که آموزش آغاز شود؛ این کار سبب می شود که اولی به عملکرد تشخیص بهتری دست یابد.

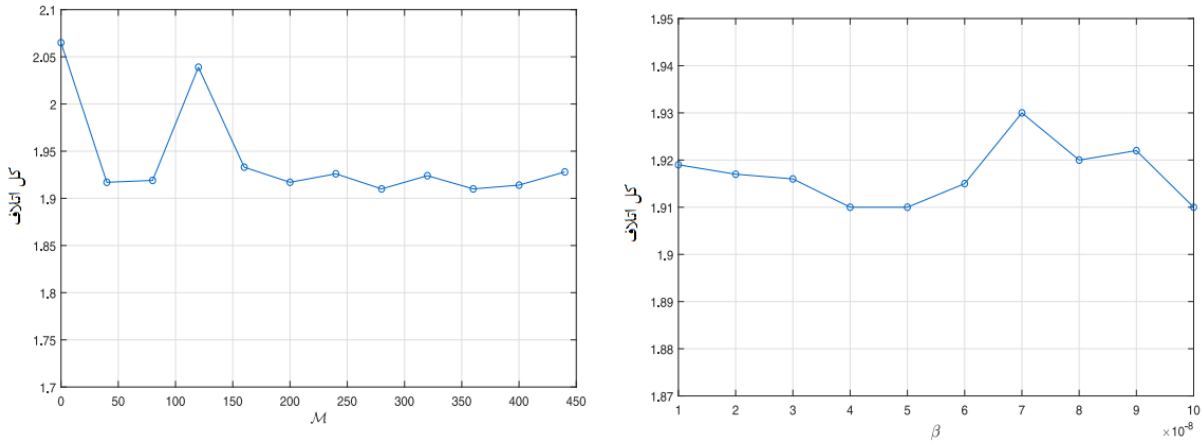
تنظیمات تست: در طی مرحله تست، ما سعی می کنم تا تنظیمات پارامتری را بیابیم که منجر به بهترین عملکرد می شوند. پارامترهای α و β در معادله (12) به ترتیب برابر مقادیر 5e-5 و 5e-8 قرار داده می شوند. حداقل حاشیه تابع MML برابر 280 تعیین می شود. ویژگی عمیق هر تصویر از خروجی لایه کاملاً متصل به دست می آید و ویژگی های هر تصویر تست اصلی و تصویر افقی چرخیده آن را به یکدیگر متصل می کنیم، بنابراین اندازه ویژگی حاصل برابر 2 در 512 بعد خواهد بود. نتایج تایید نهایی با مقایسه آستانه با فاصله اقلیدسی این دو ویژگی به دست می آید.

2-3 تحلیل تاثیر بر پارامترهای β و M

β یک فرا پارامتر برای تنظیم تاثیر MML در ترکیب است. M حداقل حاشیه طراحی شده است. این دو پارامتر بر عملکرد روش پیشنهادی تاثیرگذار هستند. بنابراین، نحوه تنظیم این دو پارامتر مورد پرسش است.

کل هزینه تنها عملکرد مدل براساس مجموعه آموزش را منعکس می کند. ما دو آزمایش را بر روی مجموعه داده VGGFace2 انجام می دهیم و تاثیر این دو پارامتر

را بر کل هزینه مورد ارزیابی قرار می دهیم. در آزمایش اول، مقدار β را برابر 5e-8 قرار می دهیم و تاثیر M بر کل هزینه را مطابق با شکل a–2 مشاهده می کنیم. در آزمایش دوم، M را برابر 280 ثابت نگه می داریم و رابطه بین β و کل هزینه را مطابق با شکل b–2 مورد ارزیابی قرار می دهیم. مطابق با شکل a–2 مشاهده می کنیم که تنظیم M به مقدار صفر و به ویژه بدون استفاده از MML صحیح نیست چرا که منجر به هزینه زیادی می شود. کمترین هزینه کل زمانی پدیدار می شود که M برابر 280 قرار داده می شود. براساس شکل b–2 مشاهده می کنیم که کل هزینه با وجود محدوده گسترده β ثابت باقی می ماند این به کمترین مقدار خود می رسد هنگامی که β برابر 5e-8 است. بنابراین، در آزمایش بعدی ما M و β را به ترتیب در مقادیر 280 و 5e-8 ثابت نگه می داریم.



شکل 2: دقت تایید چهره براساس مجموعه داده LFW همراه با دو گروه مدل: (a) β = 5e–8 ثابت و M متفاوت و (b) M=280 ثابت و β متفاوت.

3-3 چالش اول **MegaFace** در مجموعه داده **FaceScrub**

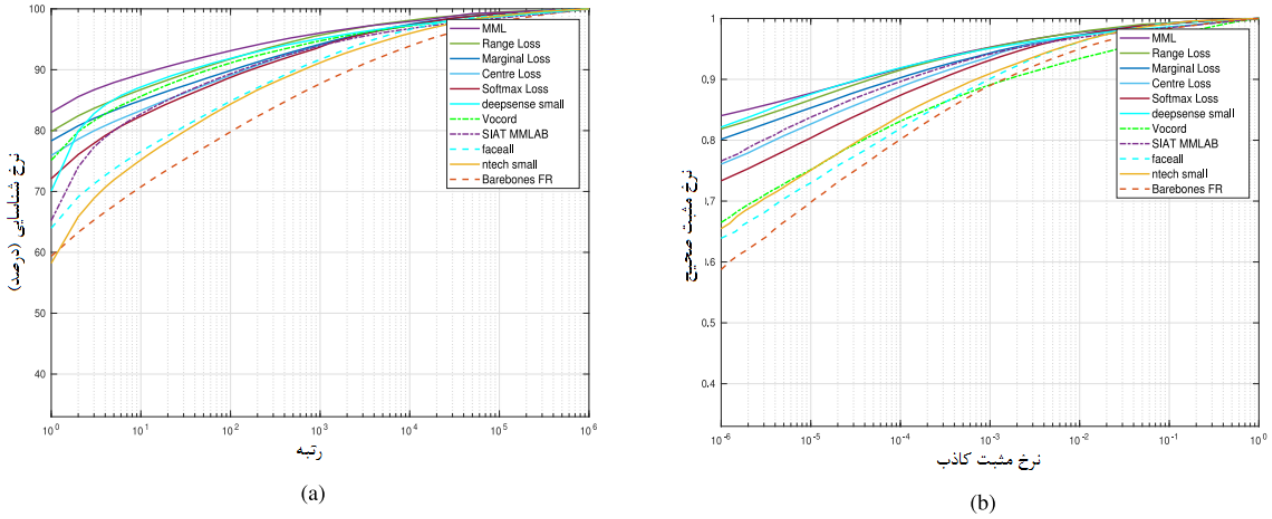
در این بخش آزمایشی را با استفاده از مجموعه داده های MegaFace [7] و FaceScrub [23] انجام می دهیم. مجموعه داده MegaFace شامل یک میلیون چهره و محدوده های مربوطه آن ها است که از مجموعه داده یاهو به نام Flickr به دست آمده است. مجموعه داده FaceScrub یک مجموعه داده عمومی شامل 0.1M تصویر از 530 هویت است. مطابق با پروتکل آزمایش MegaFace Challenge 1، مجموعه داده MegaFace به عنوان یک مجموعه distractor مورد استفاده قرار می گیرد، در حالی که مجموعه داده FaceScrub به عنوان یک مجموعه تست مورد استفاده قرار می گیرد. ارزیابی با استفاده از کد ارائه شده انجام می شود [7]. جزئیات بیشتر در رابطه با پروتکل آزمایش در مرجع [7] وجود دارد.

ما روش پیشنهادی (MML) را با تلفات مختلف و برخی روش های مبتنی بر یادگیری عمیق ارائه شده توسط تیم MegaFace مقایسه می کنیم. در آزمایش های مربوط به شناسایی چهره، منحنی های مربوط به مشخصات تطبیق انباشته[12] (CMC) [29] برای اندازه گیری قابلیت های رتبه بندی روش های مختلف محاسبه می شوند که در شکل a–3 نشان داده شده است. در آزمایش های مربوط به تایید چهره، از منحنی های مشخصات عملیاتی گیرنده[13] (ROC) برای ارزیابی روش های مختلف استفاده می کنیم. منحنی های ROC نرخ پذیرش کاذب[14] (FAR) مربوط به یک تطبیق دهنده 1:1 را برحسب نرخ عدم پذیرش کاذب[15] (FRR) تطبیق دهنده را ترسیم می کند که در

12 Cumulative Match Characteristics
13 Receiver Operating Characteristic
14 False Accept Rate
15 False Reject Rate

شکل 3–b نشان داده شده است. جدول 2 نتایج عددی روش های مختلف براساس نرخ شناسایی و نرخ تایید همراه با distractor های 1M را ارائه می کند.

طبق شکل 3–a و 3–b و همچنین جدول 2 مشاده می کنیم که روش MML عملکرد بهتری در مقایسه با روش های دیگر مبتنی بر یادگیری عمیق در تست شناسایی و تایید دارد. این موضوع اثربخشی کل ساختار را نشان می دهد. سازگاری روش MM پیشنهادی عملکرد بهتری نسبت به Marginal Loss، Center Loss، SoftMax و Range Loss دارد که اثربخشی تابع هزینه پیشنهادی را تایید می کند.



(a) (b)

شکل 3: (a) گزارش منحنی های CMC مربوط به distractor های 1M در مجموعه داده MegaFace و (b) گزارش منحنی های ROC مربوط به روش های مختلف distractorهای 1M در مجموعه داده MegaFace.

جدول 2: نرخ شناسایی و نرخ تایید روش های مختلف براساس مجموعه داده های Megaface و FaceScrub همراه با distractorهای 1M.

روش ها	Rank1@10^6	Rank100@10^6	VR @FAR10^{-6}	VR @FAR 10^{-5}
Barebones FR	59.36%	79.79%	58.77%	69.80%
ntech small	58.21%	84.34%	65.48%	75.07%
faceall	63.97%	84.84%	63.89%	72.99%
SIAT MMLAB	65.23%	89.33%	76.56%	83.78%
Vocord	75.13%	91.11%	66.50%	75.15%
deepsense small	70.06%	91.85%	82.15%	87.56%
Softmax Loss	72.11%	88.73%	73.33%	80.37%
Centre Loss	75.93%	89.07%	76.07%	82.66%
Marginal Loss	78.32%	89.87%	80.16%	85.32%
Range Loss	79.86%	91.76%	81.85%	86.65%
MML	**83.00%**	**93.12%**	**84.03%**	**87.73%**

4-3 مقایسه با بهترین روش ها در مجموعه داده های **LFW** و **YTF**

در این بخش روش پیشنهادی را براساس دو مجموعه داده عمومی یعنی LFW [20] و YTF [22] مطابق با تنظیمات ارائه شده در بخش 3-1 را مورد ارزیابی قرار می دهیم. برخ نمونه های از قبل پردازش شده مربوط به این دو مجموعه داده در شکل 4 نشان داده شده است.

شکل 4: برخی نمونه ها مربوط به مجموعه داده LFW (سمت چپ) و مجموعه داده YTF (سمت راست).

مجموعه داده LFW از طریق وب جمع آوری می شود که شامل 13233 تصویر چهره با تغییرات زیاد در پیشانی صورت، طرح و بیان است. این تصاویر چهره به 5749 هویت

مختلف مربوط می شود که 4069 مورد از آن ها دارای یک تصویر و مابقی دارای 1680 هویت با حداقل دو تصویر هستند. مجموعه داده LFW تشخیص گر چهره به نام Viola-Jones را به کار می گیرد که تنها محدودیت بر روی تصاویر جمع آوری شده است. ما از پروتکل تجربی استاندارد همراه با مجموعه داده برچسب گذاری شده [36] استفاده می کنیم و 6000 زوج چهره را مطابق با فهرست داده شده آزمایش می کنیم. مجموعه داده YTF شامل 3425 ویدیو به دست آمده از یوتیوب است. این ویدیوها به 1595 هویت با متوسط 2.15 ویدیو برای هر فرد مربوط می شود. تعداد فریم هر کلیپ ویدیویی بین 48 تا 6070 متغیر است و به طور متوسط 181.3 فریم وجود دارد. همچنین ما از پروتکل تجربی استاندارد بدون محدودیت همراه با برچسب های داده برای ارزیابی عملکرد روش های مرتبط بر روی 5000 زوج ویدیو استفاده می کنیم.

جدول 3 نتایج حاصل از روش پیشنهادی و بهترین روش های موجود در رابطه با مجموعه داده های LFW و YTF را نشان می دهد که براساس آن موارد ذیل مشاهده می شود:

- روش MML پیشنهادی عملکرد بهتری نسبت به Softmax Loss و Center Loss دارد و عملکرد براساس مجموعه داده های LFW و YTF را افزایش می دهد. در مجموعه داده LFW، دقت از 99.43 و 99.50 درصد به 99.63 درصد بهبود می یابد، در حالی که دقت در مجموعه داده YTF از 94.9 و 95.1 درصد به 95.5 درصد افزایش می یابد. همچنین روش MML عملکرد بهتری نسبت به روش های Rang Loss و Marginal Loss در مجموعه داده های LFW و YTF دارد. در مجموعه داده LFW، دقت از 99.50 و 99.52 درصد به 99.63 درصد بهبود می یابد، در حالی که دقت در

مجموعه داده YTF از 95.1 و 95.3 درصد به 95.5 درصد افزایش می یابد. این مساله اثربخشی روش MML را نشان می دهد و همچنین اثربخشی ترکیب روش های Softmax Loss، Center Loss و MML را نیز نشان می دهد.

- در مقایسه با بهترین روش ها، روش پیشنهادی دارای دقت 99.63 درصد در مجموعه داده LFW و دقت 95.5 درصد در مجموعه داده YTF است که بیشتر از روش های دیگر است. FaceNet از مجموعه داده بزرگی استفاده می کند که شامل تقریباً 200 میلیون تصویر چهره است. FaceNet نیازمند زمان بسیار بیشتری برای آموزش در مقایسه با روش پیشنهادی است که تنها از 3.05 میلیون تصویر چهره استفاده می کند.

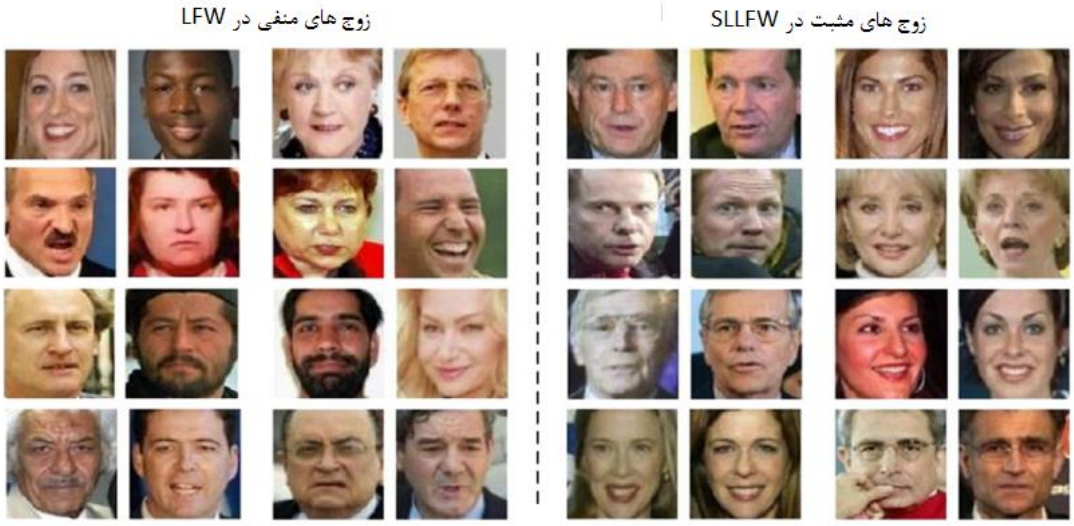
جدول 3: نرخ تایید بهترین روش ها براساس مجموعه داده های LFW و YTF.

روش ها	تصاویر	VR on LFW(%)	VR on YTF(%)
ICCV17' Range Loss [9]	1.5M	99.52	93.7
CVPR17' Marginal Loss [13]	4M	99.48	96.0
BMVC15' VGG Face [30]	2.6M	98.95	97.3
CVPR14' Deep Face [31]	4M	97.35	91.4
ICCV15' FaceNet [11]	200M	99.63	95.1
ECCV16' Centre Loss [12]	0.7M	99.28	94.9
NIPS16' Multibatch [32]	2.6M	98.20	
ECCV16' Aug [33]	0.5M	98.06	
CVPR17' SphereFace [16]	0.5M	99.42	95.0
ECCV18' Contrastive CNN [34]	0.5M	99.12	
ECCV18' OE-CNNs [35]	1.7M	99.47	
Softmax Loss	3.05M	99.43	94.9
Centre Loss	3.05M	99.50	95.1
Range Loss	3.05M	99.50	95.1
Marginal Loss	3.05M	99.52	95.3
MML (Proposed)	3.05M	99.63	95.5

5-3 مقایسه بیشتر در مجموعه داده SLLFW

از آنجا که روش های بیشتر و بیشتری به طور تدریجی حد نظری LFW را بررسی می کنند، شکاف های موجود بین روش های مختلف کمتر و کمتر می شود و تمایز بین آن ها را دشوار می کند. بنابراین، برای تایید عملکرد روش MML، آزمایش اضافی بر

روی [21] SLLFW انجام می شود. مجموعه داده SLLFW از همان زوج های مثبت به عنوان LFW برای تست استفاده می کند، اما در SLLFW تعداد 3000 زوج چهره مشابه به طور آزادانه از LFW انتخاب می شوند تا جایگزین زوج های منفی تصادفی در LFW شوند. برخی از نمونه های زوج های منفی در LFW و SLLFW در شکل 5 نشان داده شده است. در مقایسه با LFW، SLLFW چالش های بیشتری را به تست اضافه می کند و سبب می شود که دقت بهترین روش ها به میزان 10 تا 20 درصد افت کند.



شکل 5: نمونه هایی از زوج های منفی در مجموعه داده های LFW و SLLFW. در مقایسه با زوج های منفی در LFW، تشخیص زوج های منفی در SLLFW نسبتاً دشوار است.

جدول 4 دقت تایید روش های مختلف در مجموعه داده SLLFW را نشان می دهد. نتایج حاصل از برخی های معیار در نیمه بالایی این جدول نشان داده شده است. این نتایج به طور عمومی در دسترس قرار دارد [38] و توسط تیم SLLFW ارائه شده است [21]. همانطور که از جدول 4 دیده می شود، روش MML به عملکرد بهتری نسبت به

روش های معیار در SLLFW دست می یابد. همچنین روش MML دقت بیشتری را نسبت به دیگر توابع هزینه مربوطه نشان می دهد. در نیمه بالایی این حدول، دقت روش های معیار تنها بین 16.75 درصد و 4.68 درصد از LFW به SLLFW افت می کند. طبق مقایسه، دقت روش MML به میزان 3.26 درصد افت می کند. نتایج به دست آمده در SLLFW عملکرد روش های پیشنهادی را بیشتر تایید می کند.

جدول 4: عملکرد تایید روش های مختلف در مجموعه داده SLLFW.

روش	تصاویر	LFW(%)	SLLFW(%)
Deep Face [31]	0.5M	92.87	78.78
DeepID2 [10]	0.2M	95.00	78.25
VGG Face [30]	2.6M	96.70	85.78
DCMN [21]	0.5M	98.03	91.00
Noisy Softmax [37]	0.5M	99.18	94.50
Softmax Loss	3.05M	99.43	95.92
Centre Loss	3.05M	99.50	96.02
Range Loss	3.05M	99.50	96.07
Marginal Loss	3.05M	99.52	96.07
MML	3.05M	99.63	96.37

6-3 نتایج براساس مجموعه داده های IJB-B و IJB-C

مجموعه داده IJB-B [24] شامل 21.8K تصویر و 55K فریم از 7011 ویدیو است. در مجموعه داده IJB-B، تعداد 1845 موضوع وجود دارد که هیچ گونه هم پوشانی با معیارهای تشخیص چهره همانند VGGFace2 [6] و CASIA WebFace [8] ندارند. در مجموعه داده IJB-B، به طور کلی 12115 قالب همراه با 10270 تطبیق اصلی و 8M تطبیق جعلی وجود دارد. مجموعه داده IJB-C [25] توسعه ای از IJB-B است. این مجموعه شامل 31.3K تصویر ثابت و 117.5K فریم از 1179 ویدیو است. تمامی این تصاویر و ویدیوها مربوط به 3531 موضوع هستند که هیچ گونه هم پوشانی با معیارهای معروف تشخیص چهره ندارند. در مجموعه داده IJB-C، به طور کلی تعداد

23124 قالب وجود دارد که شامل 19557 تطبیق اصلی و 15693K تطبیق جعلی است.

با پیروی از پروتکل تایید 1:1، ما روش MML پیشنهادی را با جدیدترین روش های مطابق با جدول 5 مقایسه می کنیم. برای مقایسه بهتر، روش MML را به طور مستقیم با توابع هزینه مرتبط در شرایط یکسان مقایسه می کنیم. نتایج به دست آمده نشان می دهد که روش MML عملکرد بهتری نسبت به جدیدترین روش ها دارد که در بخش بالایی جدول 5 در مجموعه داده های IJB-B و IJB-C نشان داده شده است. همچنین روش MML عملکرد بهتری نسبت به توابع هزینه مربوطه در مقایسه با بخش پایینی جدول 5 نشان می دهد.

جدول 5: نتایج ارزیابی با استفاده از پروتکل تایید 1:1 در مجموعه داده های IJB-B و IJB-C.

روش	IJB-B TAR@FAR=1e−4	IJB-C TAR@FAR=1e−4
Crystal Loss [39]	0.898	0.919
ResNet50 [6]	0.784	0.825
SENet50 [6]	0.800	0.840
ResNet50+SENet50 [6]	0.800	0.841
MN-v [40]	0.818	0.852
MN-vc [40]	0.831	0.862
ResNet50+DCN(Kpts) [41]	0.850	0.867
ResNet50+DCN(Divs) [41]	0.841	0.880
SENet50+DCN(Kpts) [41]	0.846	0.874
SENet50+DCN(Divs) [41]	0.849	0.885
GAN+ArcFace [42]	0.904	0.926
PCP+ArcFace [42]	0.901	0.924
PCPSM+ArcFace [42]	0.907	0.928
LRR+ArcFace [42]	0.909	0.931
PCPSFM+ArcFace [42]	0.911	0.934
Softmax Loss	0.908	0.931
Centre Loss	0.910	0.934
Range Loss	0.916	0.937
Marginal Loss	0.917	0.939
MML	**0.921**	**0.943**

4. نتیجه گیری

در این مقاله یک تابع هزینه جدید به حداقل هزینه حاشیه ای (MML) برای هدایت شبکه های عصبی عمیق به منظور یادگیری ویژگی های چهره ارائه شده است. براساس اطلاعات موجود، روش MML اولین تابع هزینه است که تعیین حداقل حاشیه بین دسته های مختلف را در نظر می گیرد. ما نشان می دهیم که پیاده سازی تابع هزینه پیشنهادی در شبکه های CNN بسیار ساده است و مدل های CNN ما به طور مستقیم توسط SGD استاندارد قابل بهینه سازی است. آزمایش های گسترده ای بر روی هفت مجموعه داده عمومی موجود انجام شده است. ما روش MML را با روش های منتشر شده در چند سال اخیر مقایسه می کنیم. همچنین روش MML را به طور مستقیم با توابع هزینه مربوطه تحت یک ساختار معین مقایسه می کنیم. نتایج به دست آمده نشان می دهد که روش MML دارای بهترین عملکرد است. تحقیقات آتی برای تعیین خودکار حداقل حاشیه M مورد نیاز است. همچنین سعی می کنیم تا اثبات نظری مزیت تعیین حداقل حاشیه را مورد بررسی قرار دهیم.

References

[1] J.M. Pandya, D. Rathod, J.J. Jadav, A survey of face recognition approach, Int. J. Eng. Res. Appl. (IJERA) 3 (1) (2013) 632–635.

[2] L. Wu, Y. Wang, J. Gao, X. Li, Deep adaptive feature embedding with local sample distributions for person re-identification, Pattern Recognit. 73 (2018) 275–288.

[3] J. Han, D. Zhang, G. Cheng, N. Liu, D. Xu, Advanced deep-learning techniques for salient and category-specific object detection: a survey, IEEE Signal Process. Mag. 35 (1) (2018) 84–100.

[4] M. Ma, N. Marturi, Y. Li, A. Leonardis, R. Stolkin, Region-sequence based six-stream CNN features for general and fine-grained human action recognition in videos, Pattern Recognit. 76 (2018) 506–521.

[5] Y. Guo, L. Zhang, Y. Hu, X. He, J. Gao, MS-Celeb-1M: a dataset and benchmark for large scale face recognition, in: European Conference on Computer Vision, Springer, 2016.

[6] Q. Cao, L. Shen, W. Xie, O.M. Parkhi, A. Zisserman, Vggface2: a dataset for recognising faces across pose and age, in: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE, 2018, pp. 67–74.

[7] I. Kemelmacher-Shlizerman, S.M. Seitz, D. Miller, E. Brossard, The megaface benchmark: 1 million faces for recognition at scale, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4873–4882.

[8] D. Yi, Z. Lei, S. Liao, S.Z. Li, Learning face representation from scratch, 28 Nov 2014, arXiv:1411.7923v1 (2014).

[9] X. Zhang, Z. Fang, Y. Wen, Z. Li, Y. Qiao, Range loss for deep face recognition with long-tailed training data, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 5419–5428.

[10] Y. Sun, Y. Chen, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, in: Advances in neural information processing systems, 2014, pp. 1988–1996.

[11] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: a unified embedding for face recognition and clustering, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 815–823.

[12] Y. Wen, K. Zhang, Z. Li, Y. Qiao, A discriminative feature learning approach for deep face recognition, in: Computer Vision – ECCV 2016, Lecture Notes in Computer Science, Springer, Cham, 2016, pp. 499–515.

[13] J. Deng, Y. Zhou, S. Zafeiriou, Marginal loss for deep face recognition, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 2006–2014.

[14] R. Ranjan, C.D. Castillo, R. Chellappa, L2-constrained softmax loss for discriminative face verification, 7 Jun 2017, arXiv:1703.09507v3 (2017).

[15] W. Liu, Y. Wen, Z. Yu, M. Yang, Large-margin softmax loss for convolutional neural networks, in: International Conference on Machine Learning, 2016, pp. 507–516.

[16] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, L. Song, Sphereface: deep hypersphere embedding for face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 212–220.

[17] F. Wang, J. Cheng, W. Liu, H. Liu, Additive margin softmax for face verification, IEEE Signal Process. Lett. 25 (7) (2018) 926–930.

[18] J. Deng, J. Guo, S. Zafeiriou, Arcface: additive angular margin loss for deep face recognition, 9 Feb 2019, arXiv:1801.07698v3 (2018).

[19] Y. Zheng, D.K. Pal, M. Savvides, Ring loss: Convex feature normalization for face recognition, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[20] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: a Database for Studying Face Recognition in Unconstrained Environments, Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[21] W. Deng, J. Hu, N. Zhang, B. Chen, J. Guo, Fine-grained face verification: Fglfw database, baselines, and human-dcmn partnership, Pattern Recognit. 66 (2017) 63–73.

[22] L. Wolf, T. Hassner, I. Maoz, Face recognition in unconstrained videos with matched background similarity, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE, 2011, pp. 529–534.

[23] H.-W. Ng, S. Winkler, A data-driven approach to cleaning large face datasets, in: Image Processing (ICIP), 2014 IEEE International Conference on, IEEE, 2014, pp. 343–347.

[24] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, N. Kalka, A.K. Jain, J.A. Duncan, K. Allen, J. Cheney, P. Grother, Iarpa janus benchmark-b face dataset, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2017.

[25] B. Maze, J. Adams, J.A. Duncan, N. Kalka, T. Miller, C. Otto, A.K. Jain, W.T. Niggel, J. Anderson, J. Cheney, P. Grother, Iarpa janus benchmark - c: Face dataset and protocol, in: 2018 International Conference on Biometrics (ICB), 2018, pp. 158–165, doi:10.1109/ICB2018.2018.00033.

[26] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, IEEE Signal Process. Lett. 23 (10) (2016) 1499–1503.

[27] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017.

[28] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, et al., Tensorflow: large-scale machine learning on heterogeneous distributed systems, Mar 2016, arXiv:1603.04467v2 (2016).

[29] P.J. Phillips, P. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, M. Bone, R.V.T. FACE, Evaluation report, Facial Recognit. Vendor Test 2002, 2003.

[30] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition., in: 2015 British Machine Vision Conference (BMVC), 1, 2015, p. 6.

[31] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, DeepFace: closing the gap to human-level performance in face verification, in: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '14, IEEE Computer Society, Washington, DC, USA, 2014, pp. 1701–1708.

[32] O. Tadmor, T. Rosenwein, S. Shalev-Shwartz, Y. Wexler, A. Shashua, Learning a metric embedding for face recognition using the multibatch method, in: Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16, Curran Associates Inc., USA, 2016, pp. 1396–1397.

[33] I. Masi, A.T. Tran, T. Hassner, J.T. Leksut, G. Medioni, Do we really need to collect millions of faces for effective face recognition? in: Computer Vision – ECCV 2016, Lecture Notes in Computer Science, Springer, Cham, 2016, pp. 579–596.

[34] C. Han, S. Shan, M. Kan, S. Wu, X. Chen, Face recognition with contrastive convolution, in: The European Conference on Computer Vision (ECCV), 2018.

[35] Y. Wang, D. Gong, Z. Zhou, X. Ji, H. Wang, Z. Li, W. Liu, T. Zhang, Orthogonal deep features decomposition for age-invariant face recognition, in: The European Conference on Computer Vision (ECCV), 2018.

[36] G.B. Huang, E. Learned-Miller, Labeled faces in the wild: updates and new reporting procedures, Tech. Rep, 14–003, Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, 2014.

[37] B. Chen, W. Deng, J. Du, Noisy softmax: improving the generalization ability of dcnn via postponing the early softmax saturation, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[38] The results of some baseline methods provided by SLLFW team: http://www.whdeng.cn/SLLFW/#results.

[39] R. Ranjan, A. Bansal, J. Zheng, H. Xu, J. Gleason, B. Lu, A. Nanduri, J.-C. Chen, C.D. Castillo, R. Chellappa, A fast and accurate system for face detection, identification, and verification, IEEE Trans. Biom., Behav. Identity Sci. 1 (2) (2019) 82–96.

[40] W. Xie, A. Zisserman, Multicolumn networks for face recognition, 24 Jul 2018, arXiv:1807.09192v1 (2018).

[41] W. Xie, L. Shen, A. Zisserman, Comparator networks, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 782–797.

[42] N. Xue, J. Deng, S. Cheng, Y. Panagakis, S.P. Zafeiriou, Side information for face completion: a robust pca approach, IEEE Trans. Pattern Anal. Mach. Intell. (2019). 1–1, doi: 10.1109/TPAMI.2019.2902556.