

Facial expressions recognition

استاد راهنما: دکتر ستوده نژاد سروری

نام دانشجو: ملیکا یزدان پناه

MAIN PAPER

این مقاله به یکی از مشکلات اساسی در حوزه تشخیص احساسات چهره، یعنی مواجهه با دسته‌های باز (Open-Set) از احساسات، و راهکار پیشنهادی نویسنده‌ها برای رفع این چالش اشاره دارد.

مسئله: مدل‌های سنتی تشخیص احساسات چهره (FER) معمولاً بر روی مجموعه داده‌هایی آموزش داده می‌شوند که شامل هفت دسته‌بندی اساسی از احساسات چهره هستند. اما در کاربردهای دنیای واقعی، احساسات بسیاری وجود دارند که در این دسته‌بندی‌ها قرار نمی‌گیرند، از جمله احساسات ترکیبی و ناشناخته. این مسئله باعث می‌شود که مدل‌های FER در شناسایی احساسات جدید دچار خطا شوند و نتوانند به درستی با احساسات چهره ناشناخته تطابق پیدا کنند.

راه‌حل پیشنهادی: مقاله به معرفی وظیفه جدیدی به نام "تشخیص احساسات چهره در فضای باز" (open-set FER) می‌پردازد. در این روش، نویسنده‌ها چالش تشخیص احساسات ناشناخته را به یک مسئله تشخیص برچسب‌های نویزی تبدیل می‌کنند. به دلیل فاصله کم بین دسته‌های مختلف احساسات چهره، تمایز میان احساسات شناخته‌شده و ناشناخته دشوار می‌شود. برای رفع این مشکل، یک روش مبتنی بر سازگاری نقشه توجه (Attention Map Consistency) و آموزش چرخه‌ای (Cycle Training) پیشنهاد شده است. این روش از فاصله کم بین دسته‌های مختلف احساسات به عنوان یک مزیت استفاده می‌کند و به شناسایی بهتر نمونه‌های دسته باز کمک می‌کند.

نتایج: آزمایشات نشان می‌دهد که روش پیشنهادی به طور قابل توجهی بهتر از روش‌های موجود مانند DIAS و OpenMax عمل می‌کند و در تشخیص احساسات ناشناخته عملکرد بهتری دارد.

:ABSTRACT

: مدل‌های تشخیص احساسات چهره (FER) نقش مهمی در تعاملات انسان و ماشین ایفا می‌کنند و به دستگاه‌ها کمک می‌کنند تا احساسات انسان را بهتر درک کنند. با این حال، این مدل‌ها با چالش‌های بزرگی مواجه هستند، چرا که اغلب در مواجهه با احساسات چهره ناشناخته که در دسته‌بندی‌های از پیش تعریف‌شده نمی‌گنجند، ناتوان هستند.

چالش‌ها: مدل‌های فعلی اغلب احساسات جدید را به اشتباه به یکی از دسته‌های شناخته‌شده نسبت می‌دهند. یکی از دلایل اصلی این مسئله این است که مدل‌های یادگیری عمیق به‌طور معمول در مواجهه با داده‌های ناشناخته، سطح بالایی از اطمینان دارند، حتی اگر داده‌ها قبلاً دیده نشده باشند. روش‌های فعلی تشخیص دسته باز، مانند DIAS و OpenMax، در FER به خوبی عمل نمی‌کنند زیرا فاصله بین دسته‌های مختلف احساسات چهره در انسان‌ها بسیار کم است و این امر باعث می‌شود که احساسات ناشناخته بسیار شبیه به احساسات شناخته‌شده به نظر برسند.

مشارکت: نویسندگان این مقاله پیشنهاد می‌کنند که این مشکل را می‌توان به یک مزیت تبدیل کرد و با در نظر گرفتن این چالش به عنوان یک مسئله تشخیص برچسب‌های نویزی، می‌توان تشخیص بهتری از احساسات دسته باز ارائه داد. روش ارائه‌شده نه تنها به بهبود دقت تشخیص کمک می‌کند، بلکه قدرت تمایز بیشتری در مواجهه با احساسات ناشناخته دارد.

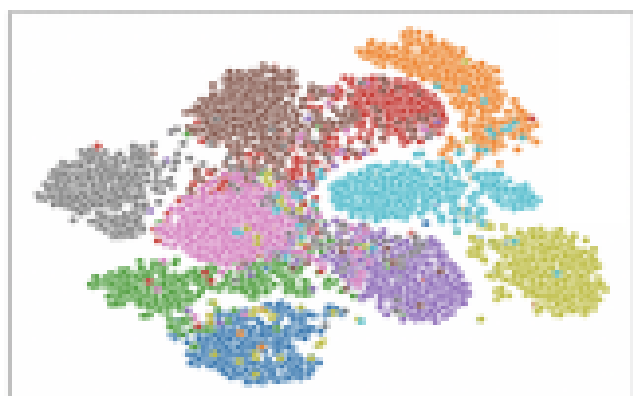
: introduction

****پیش‌زمینه:**** مدل‌های تشخیص احساسات چهره (FER) برای تعامل انسان و کامپیوتر بسیار حیاتی هستند و به ماشین‌ها کمک می‌کنند تا احساسات انسانی را تفسیر کنند. با این حال، مدل‌های فعلی در مواجهه با حالات ناشناخته چهره که در دسته‌های از پیش تعریف شده قرار نمی‌گیرند، مشکل دارند.

****چالش‌ها:**** این مدل‌ها اغلب حالات جدید را به اشتباه به عنوان حالات موجود دسته‌بندی می‌کنند، زیرا مدل‌های یادگیری عمیق در مواجهه با داده‌های دیده نشده، سطوح بالایی از اعتماد به نفس نشان می‌دهند. روش‌های فعلی تشخیص دسته‌باز (مانند DIAS و OpenMax) عملکرد مناسبی در تشخیص احساسات چهره ندارند، زیرا چهره‌های انسانی دارای فاصله کم بین دسته‌های مختلف هستند که باعث می‌شود حالات ناشناخته بسیار شبیه به حالات شناخته‌شده به نظر برسند.

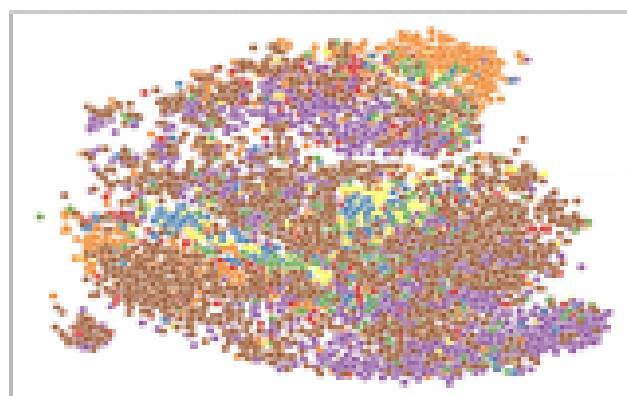
****مشارکت:**** نویسندگان پیشنهاد می‌کنند که مشکل فاصله کم بین دسته‌ها را به یک مزیت تبدیل کنند و این چالش را به عنوان یک مسئله برچسب نویزی در نظر بگیرند، که این رویکرد به تشخیص بهتر حالات ناشناخته در دسته‌باز کمک می‌کند.

Figure 1 (Extracted Features of CIFAR-10 and RAF-DB) ###



CIFAR-10

DIAS AUROC: 0.850



RAF-DB

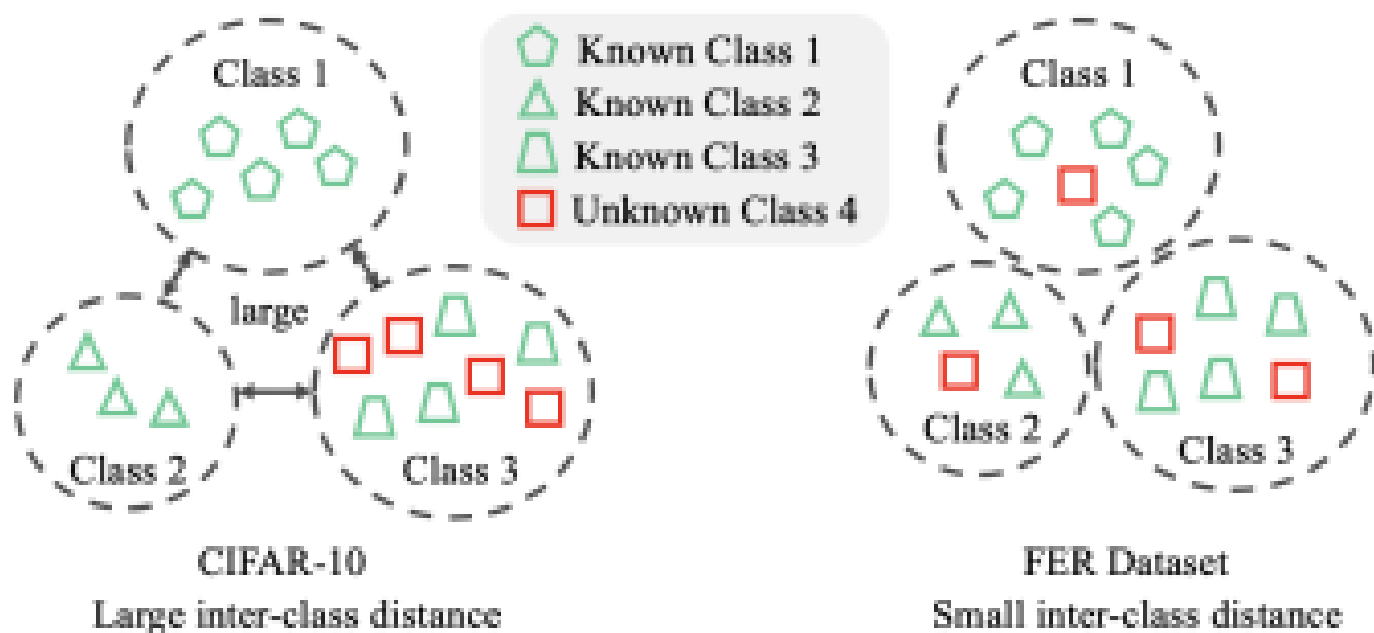
DIAS AUROC: 0.714

Method	RAF-DB	FERPlus	AffectNet
DIAS	0.714	0.709	0.551
Ours	0.909 (+27.3%)	0.882 (+24.4%)	0.674 (+22.3%)

این شکل ویژگی‌های استخراج‌شده از دو مجموعه داده: CIFAR-10 و RAF-DB را نشان می‌دهد. این شکل تفاوت بین فاصله زیاد بین کلاس‌های CIFAR-10 و فاصله کم بین کلاس‌های RAF-DB را مقایسه می‌کند.

CIFAR-10: ویژگی‌ها پراکنده‌تر هستند که باعث می‌شود تشخیص نمونه‌های مجموعه باز از نمونه‌های مجموعه بسته آسان‌تر شود. RAF-DB: ویژگی‌ها بسیار به یکدیگر نزدیک‌تر هستند، به این معنی که نمونه‌های مجموعه باز بسیار شبیه به نمونه‌های مجموعه بسته ظاهر می‌شوند. به همین دلیل روش‌های پیشرفته فعلی مانند DIAS عملکرد ضعیف‌تری در تشخیص حالت‌های چهره دارند (به عنوان مثال، AUROC از 0.850 به 0.714 کاهش می‌یابد). روش ارائه شده توسط نویسندگان بهبودهای قابل توجهی (بیش از ۲۰٪ در AUROC) در تشخیص حالات چهره در مجموعه باز به دست می‌آورد.

Figure 2 (Illustration of Noisy Label Detection) ###



این شکل نشان می‌دهد که چگونه روش نویسندگان تشخیص برچسب‌های نویزی را مدیریت می‌کند:

در مجموعه داده‌هایی مانند CIFAR-10، بیشتر نمونه‌های مجموعه باز (مثلاً از کلاس "گربه") به نزدیک‌ترین کلاس مجموعه بسته (مثلاً "سگ") اختصاص داده می‌شوند.

در تشخیص حالت‌های چهره (FER)، نمونه‌های مجموعه باز بین تمام کلاس‌های مجموعه بسته توزیع می‌شوند. شکل، توزیع برچسب‌های شبه‌ای (pseudo labels) نمونه‌های FER مجموعه باز را در کلاس‌های مختلف نشان می‌دهد. این نویز متقارن تشخیص‌پذیرتر از نویز نامتقارن است، جایی که نمونه‌های مجموعه باز بیشتر به یک کلاس معنایی مشابه تعلق می‌گیرند.

RELATED WORK – FACIAL EXPRESSION RECOGNITION : (overview of prior studies and techniques in the field of Facial Expression Recognition (FER).)

روش‌های قبلی:

روش‌های مختلفی برای FER پیشنهاد شده است، از جمله:

لی، دنگ و دو (۲۰۱۷): از جمع‌سپاری برای شبیه‌سازی شناسایی احساسات انسانی استفاده کردند و روش‌هایی را معرفی کردند که سعی دارند نحوه شناسایی احساسات توسط انسان‌ها را منعکس کنند.

فرزانه و کی (۲۰۲۱): واریانت جدیدی از مرکزیت را پیشنهاد کردند تا هم شباهت درون‌کلاسی (اطمینان از اینکه تصاویر از همان کلاس نزدیک به هم هستند) و هم جداسازی بین‌کلاسی (نگه‌داشتن تصاویر کلاس‌های مختلف دور از هم) را به حداکثر برسانند.

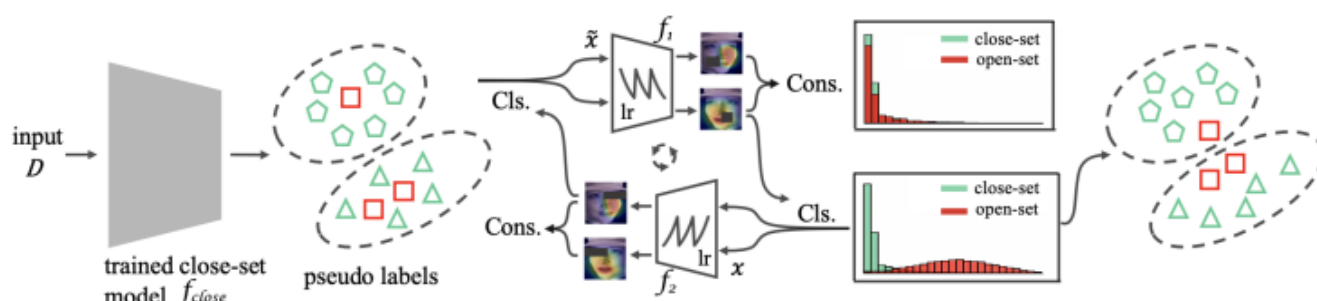
ژانگ، وانگ و دنگ (۲۰۲۱): یادگیری مقایسه‌ای نسبی را معرفی کردند که بر آموزش مدل‌های FER از طریق مقایسه‌ی شباهت یا تفاوت حالات چهره تمرکز دارد.

رؤان و همکاران (۲۰۲۱): روش‌هایی برای استخراج ویژگی‌های مرتبط با احساسات از طریق تجزیه طراحی کردند و بر مهم‌ترین جنبه‌های حالات چهره تمرکز کردند.

محدودیت‌های روش‌های موجود:

این پاراگراف به این نکته اشاره دارد که در حالی که این روش‌ها در FER بسته (جایی که تعداد کلاس‌ها ثابت است) مؤثر هستند، اما برای مدیریت FER باز ناکافی هستند. بیشتر روش‌های FER موجود پیش‌بینی‌های مطمئن برای داده‌های باز (احساسات ناشناخته) انجام می‌دهند که محدودیت‌هایی در کاربردهای دنیای واقعی ایجاد می‌کند، زیرا نمی‌توانند احساساتی را که بخشی از مجموعه داده‌های آموزشی نبوده‌اند، به درستی شناسایی کنند.

Pipeline of the Proposed Method



این شکل روند روش پیشنهادی تشخیص حالت‌های چهره در مجموعه باز (open-set FER) را ارائه می‌دهد. این روند شامل مراحل زیر است:

استفاده از یک مدل از پیش آموزش‌دیده شده برای مجموعه بسته جهت تولید برچسب‌های شبه‌ای (pseudo labels) برای هر دو نمونه مجموعه بسته و مجموعه باز.

آموزش دو مدل FER به صورت چرخشی با استفاده از زیان سازگاری نقشه توجه (attention map consistency loss) و نرخ یادگیری چرخشی.

نرخ یادگیری چرخشی یک مجموعه از مدل‌ها ایجاد می‌کند و سازگاری نقشه توجه مانع از حفظ برچسب‌های نویزی مجموعه بسته توسط مدل می‌شود.

پس از آموزش، نمونه‌های مجموعه باز دارای زیان طبقه‌بندی بالا و نمونه‌های مجموعه بسته دارای زیان طبقه‌بندی کم هستند که به مدل‌ها اجازه می‌دهد نمونه‌های مجموعه باز را از نمونه‌های مجموعه بسته تفکیک کنند.

open set recognition

شناسایی باز (Open-Set Recognition)

این بخش به چالش شناسایی باز (OSR) می‌پردازد، که زمانی اتفاق می‌افتد که مدل‌ها با کلاس‌هایی مواجه می‌شوند که در طول آموزش مشاهده نشده‌اند. دو جریان اصلی از روش‌های OSR وجود دارد:

مدل‌های تفکیکی: این روش‌ها از دسته‌بندها برای تشخیص بین کلاس‌های شناخته‌شده و ناشناخته استفاده می‌کنند (به‌عنوان مثال، OpenMax). این مدل‌ها از استراتژی‌هایی مانند جایگزینی لایه softmax با روش‌های کالیبراسیون توزیع دیگر استفاده می‌کنند.

مدل‌های تولیدی: این مدل‌ها توزیع کلاس‌های نادیده را با تولید داده‌هایی که به آن‌ها نزدیک است، پیش‌بینی می‌کنند. این رویکردها به تکنیک‌هایی مانند شبکه‌های رقابتی مولد (GAN) وابسته هستند.

با این حال، هر دو جریان در datasets که فاصله‌های بین کلاس‌ها بزرگ است (مانند CIFAR-10) بهتر عمل می‌کنند، اما زمانی که بر روی داده‌های FER که در آن‌ها احساسات چهره ظریف است و تفاوت‌های بین کلاس‌ها کوچکتر است، اعمال شوند، شکست می‌خورند. این بخش تأکید می‌کند که روش‌های موجود OSR برای FER مناسب نیستند

problem definition

تعریف مشکل

نویسندگان مشکل شناسایی احساسات چهره در حالت باز (Open-Set FER) را به صورت زیر تعریف می‌کنند:

شناسایی احساسات چهره در حالت بسته (Closed-set FER): بر روی داده‌هایی با تعداد مشخصی از کلاس‌ها (مثلاً هفت احساس اصلی) آموزش می‌بیند.

شناسایی احساسات چهره در حالت باز (Open-set FER): با ابرازهای جدید و قبلاً دیده‌نشده در یک محیط واقعی مواجه می‌شود. چالش این است که مشخص شود آیا یک نمونه خاص به یک کلاس شناخته‌شده تعلق دارد یا از یک کلاس نادیده (کلاس $K+1$ ، کلاس جدید) است.

مشکل شناسایی احساسات چهره این است که نمونه‌های حالت باز (open-set) بسیار مشابه نمونه‌های حالت بسته (closed-set) به نظر می‌رسند به دلیل فاصله کوچک بین کلاس‌ها، که منجر به اعتماد به نفس بیش از حد در مدل‌های یادگیری عمیق می‌شود که ابرازهای دیده‌نشده را به یکی از کلاس‌های شناخته‌شده طبقه‌بندی می‌کنند. یک روش جدید برای مقابله با این چالش ضروری است.

METHOD

pseudo labels

یک مدل FER در حالت بسته آموزش داده می‌شود و برای تمامی نمونه‌ها برچسب‌های شبه تولید می‌کند. نمونه‌های حالت باز، برچسب‌های نادرست به تمام کلاس‌های شناخته‌شده اختصاص می‌گیرند که در واقع برچسب‌های "نویزی" ایجاد می‌کند.

برخلاف مجموعه داده‌های سنتی، در FER برچسب‌های نادرست بین چندین کلاس شناخته‌شده توزیع می‌شوند، به جای اینکه به یک کلاس معنایی مشابه متمرکز شوند.

cycle trainig :

دو مدل به صورت چرخه‌ای آموزش داده می‌شوند. یک مدل با برچسب‌های شبه آموزش می‌بیند و نمونه‌های پاک را برای آموزش مدل دیگر انتخاب می‌کند. این مدل‌ها به طور متناوب یکدیگر را آموزش می‌دهند و یاد می‌گیرند.

attention map consistency :

برای جلوگیری از حفظ برچسب‌های نادرست توسط مدل، از ثبات نقشه توجه استفاده می‌شود. این تکنیک اطمینان می‌دهد که مدل روی ویژگی‌های مهم چهره تمرکز دارد، حتی زمانی که برچسب‌های شبه نویزی هستند.

cycling learning rate :

یک نرخ یادگیری چرخه‌ای استفاده می‌شود تا از بیش‌برازش جلوگیری شود. این کار با شبیه‌سازی یک مجموعه از مدل‌ها در مراحل مختلف یادگیری انجام می‌شود. زیان طبقه‌بندی نمونه‌های حالت باز زیاد خواهد بود و این امر به مدل اجازه می‌دهد تا نمونه‌های حالت باز (با زیان بالا) را از نمونه‌های حالت بسته (با زیان کم) جدا کند.

experiments

تنظیمات تجربی و نتایج برای نشان دادن اثربخشی روش آنها در چندین مجموعه داده FER ارائه شده است. مجموعه داده‌های استفاده شده:

RAF-DB: یک مجموعه داده پرکاربرد برای FER با هفت کلاس بیان اساسی.

FERplus: یک توسعه از مجموعه داده FER2013 با برچسب‌های بیشتر برای بهبود عملکرد.

AffectNet: یک مجموعه داده بزرگ مقیاس با بیش از 280,000 تصویر آموزشی در هشت کلاس.

شاخص‌ها:

عملکرد روش‌های تشخیص فضای باز (Open-Set Recognition) با استفاده از دو معیار کلیدی ارزیابی می‌شود:

AUROC: مساحت زیر منحنی ویژگی عملیاتی گیرنده. مقدار بالاتر نشان‌دهنده عملکرد بهتر در تشخیص فضای باز است.

FPR@TPR95: نرخ مثبت کاذب زمانی که نرخ مثبت واقعی 95٪ است. مقادیر کمتر نشان‌دهنده عملکرد بهتر است.

نتایج:

روش پیشنهادی در مقایسه با روش‌های OSR موجود مانند DIAS، EOW و PROSER در همه مجموعه داده‌ها به‌طور قابل توجهی بهتر عمل می‌کند، و در بسیاری از موارد بیش از 20٪ بهبود AUROC دارد.

این روش همچنین در هنگام آزمایش بر روی بیان‌های ترکیبی (که شامل ترکیبی از بیان‌های اساسی هستند) و در مجموعه داده‌های بزرگ‌تری مانند AffectNet عملکرد خوبی نشان می‌دهد، حتی با وجود اینکه این مجموعه داده‌ها نویز بیشتری دارند.

تشخیص آنلاین:

یکی از نتایج جالب این است که این روش را می‌توان در محیط آنلاین نیز اعمال کرد، جایی که تشخیص برای نمونه‌های آزمون فردی بدون نیاز به بازآموزی مدل انجام می‌شود. کاهش عملکرد تنها حدود 2.6٪ است، که همچنان بهتر از سایر روش‌های پیشرفته OSR است.

نکات کلیدی از این بخش‌ها:

مسئله Open-Set FER به دلیل فاصله کم بین کلاس‌ها در داده‌های بیان چهره چالش‌برانگیز است.

نویسندگان پیشنهاد می‌کنند که این مشکل را به مسئله تشخیص برچسب نویزی تبدیل کنند، با استفاده از برچسب‌های شبه و نقشه‌های توجه برای تشخیص زمانی که یک بیان از یک کلاس نادیده گرفته شده است.

آموزش چرخه‌ای و نرخ یادگیری چرخه‌ای از بیش‌برازش جلوگیری کرده و با شبیه‌سازی یک مجموعه از مدل‌ها عملکرد مدل را بهبود می‌بخشد.

نتایج آزمایش‌ها تأیید می‌کند که این رویکرد به‌طور قابل توجهی تشخیص فضای باز را در FER بهبود می‌بخشد و از روش‌های فعلی پیشی می‌گیرد.

novelty and contribution

نوآوری و مشارکت‌ها

این بخش به مشارکت‌های جدید مقاله می‌پردازد:

نویسندگان برای اولین بار مسئله Open-Set FER را پیشنهاد می‌دهند و به محدودیت‌های مدل‌های FER موجود هنگام مواجهه با کلاس‌های ناشناخته (بیان‌هایی که در مجموعه آموزشی نیستند) پرداخته‌اند.

آنها Open-Set FER را به مسئله تشخیص برچسب‌های نویزی تبدیل کرده‌اند، با بهره‌گیری از فاصله کم بین کلاس‌ها در داده‌های بیان چهره، که روش‌های موجود تشخیص فضای باز با آن مشکل دارند.

این رویکرد جدید از آموزش چرخه‌ای و سازگاری نقشه‌های توجه استفاده می‌کند تا تفاوت بین نمونه‌های فضای باز (ناشناخته) و بسته (شناخته) را بهتر تشخیص دهد.

روش آنها در مقایسه با روش‌های پیشرفته تشخیص فضای باز با اختلاف زیاد بهتر عمل می‌کند و اثربخشی رویکرد آنها را نشان می‌دهد.

implementation

این بخش به تنظیمات تجربی می‌پردازد:

مجموعه داده‌ها: از مجموعه داده‌های RAF-DB، FERPlus و AffectNet برای ارزیابی روش Open-Set FER استفاده می‌شود.

معماری شبکه: یک ResNet-18 برای آموزش مدل‌های FER استفاده شده است.

تنظیمات آموزشی:

نرخ یادگیری (η) به 0.0002 تنظیم شده و بهینه‌ساز Adam با وزن کاهش‌یافته 0.0001 استفاده شده است.

آموزش به مدت 40 دوره انجام شده است.

نویسندگان از معیارهای AUROC و FPR@TPR95 برای ارزیابی استفاده کرده‌اند و تمرکز آنها بر جداسازی نمونه‌های فضای باز از بسته بوده است.

آنها تأکید می‌کنند که روش آنها دقت طبقه‌بندی فضای بسته را تحت تأثیر قرار نمی‌دهد، بلکه برای بهبود تشخیص فضای باز طراحی شده است.

open-set FER with one or several basic classes

Open-Set FER با یک یا چند کلاس اساسی

این بخش به عملکرد مدل هنگام مواجهه با یک یا چند کلاس فضای باز می‌پردازد. نکات کلیدی:

جدول 1 نشان می‌دهد که روش نویسندگان بالاترین AUROC را هنگامی که تنها یک کلاس فضای باز است و همچنین زمانی که چندین کلاس فضای باز هستند، به دست می‌آورد.

این روش به‌طور مداوم بهتر از روش‌هایی مانند EOW، DIAS و PROSER عمل می‌کند، به ویژه هنگامی که بیش از یک کلاس فضای باز است، که نشان‌دهنده استحکام این رویکرد است.

Metric	AUROC (\uparrow)					FPR@TPR95 (\downarrow)					Metric	AUROC (\uparrow)				
Method	Baseline	EOW	PROS	DIAS	Ours	Baseline	EOW	PROS	DIAS	Ours	Method	Baseline	EOW	PROS	DIAS	Ours
R Sur.	0.517	0.648	0.806	0.725	0.918	0.926	0.897	0.730	0.850	0.608	F Sur.	0.406	0.641	0.676	0.710	0.933
R Fea.	0.411	0.577	0.706	0.660	0.907	0.980	0.946	0.882	0.918	0.444	F Fea.	0.370	0.581	0.664	0.634	0.899
R Dis.	0.473	0.609	0.788	0.728	0.910	0.925	0.914	0.730	0.863	0.462	F Dis.	0.352	0.596	0.771	0.711	0.871
R Hap.	0.554	0.606	0.695	0.703	0.892	0.852	0.904	0.881	0.823	0.528	F Hap.	0.476	0.645	0.731	0.726	0.855
R Sad.	0.506	0.654	0.738	0.668	0.911	0.930	0.940	0.798	0.879	0.519	F Sad.	0.413	0.575	0.681	0.665	0.862
R Ang.	0.450	0.720	0.704	0.734	0.906	0.937	0.814	0.877	0.857	0.684	F Ang.	0.410	0.578	0.798	0.753	0.891
R Neu.	0.566	0.606	0.803	0.778	0.917	0.866	0.977	0.777	0.819	0.587	F Neu.	0.544	0.529	0.852	0.767	0.864
Mean	0.497	0.631	0.749	0.714	0.909	0.917	0.913	0.811	0.858	0.547	Mean	0.424	0.592	0.739	0.709	0.882

این جدول عملکرد تشخیص روش‌های پیشرفته شناسایی حالت‌های باز مانند Baseline، EOW، PROSER، DIAS و روش پیشنهادی را مقایسه می‌کند.

شاخص‌ها:

AUROC (مساحت زیر منحنی ROC): اندازه‌گیری می‌کند که مدل چقدر خوب نمونه‌های باز را از نمونه‌های بسته جدا می‌کند (هرچه بیشتر باشد بهتر است).

FPR@TPR95: نرخ مثبت کاذب وقتی که نرخ مثبت واقعی ۹۵٪ است (هرچه کمتر باشد بهتر است).

نتایج نشان می‌دهد که روش پیشنهادی به طور قابل توجهی عملکرد بهتری نسبت به سایر روش‌ها دارد، با مقادیر AUROC به اندازه ۰.۹۱۸ برای برخی کلاس‌ها (مثلاً Surprise)، در حالی که سایر روش‌ها در حدود ۰.۷ یا کمتر هستند.

Open class	Baseline	EOW	PROSER	DIAS	Ours
Sur.+Fea.	0.436	0.561	0.763	0.706	0.916
Fea.+Dis.	0.445	0.583	0.764	0.688	0.884
Dis.+Hap.	0.575	0.632	0.727	0.700	0.887
Hap.+Sad.	0.609	0.536	0.726	0.717	0.865
Sad.+Ang.	0.486	0.691	0.718	0.702	0.879
Ang.+Neu.	0.552	0.634	0.825	0.769	0.895
Sur.+Fea.+Dis.	0.497	0.588	0.769	0.731	0.893
Fea.+Dis.+Hap.	0.592	0.534	0.667	0.685	0.880
Dis.+Hap.+Sad.	0.649	0.624	0.743	0.723	0.815
Hap.+Sad.+Ang.	0.637	0.698	0.705	0.718	0.840
Sad.+Ang.+Neu.	0.630	0.750	0.829	0.802	0.883
Mean	0.555	0.621	0.749	0.722	0.876

این جدول عملکرد زمانی که چندین کلاس به صورت باز-مجموعه هستند (دو یا سه کلاس باز به طور همزمان) را نشان می‌دهد.

روش پیشنهادی باز هم بهترین عملکرد را در تمام تنظیمات نشان می‌دهد و به AUROC برابر با ۰.۸۷۶ با کلاس‌های چندگانه باز دست می‌یابد، در حالی که سایر روش‌ها عملکرد ضعیف‌تری دارند (در حدود ۰.۷).

Dataset	Baseline	EOW	PROSER	DIAS	Ours
Compound	0.665	0.679	0.648	0.674	0.771
AffectNet	0.552	0.507	0.610	0.551	0.674

این جدول عملکرد روش را بر روی حالات ترکیبی از مجموعه داده‌های RAF-DB و AffectNet ارزیابی می‌کند.

حالات ترکیبی تشخیص سخت‌تری دارند، زیرا به حالات پایه شبیه‌تر هستند.

روش پیشنهادی به AUROC برابر با ۰.۷۷۱ برای کلاس‌های ترکیبی و ۰.۶۷۴ برای AffectNet دست می‌یابد و عملکرد بهتری نسبت به سایر روش‌ها دارد، که به حدود ۰.۵ تا ۰.۶ کاهش می‌یابند.

compound classes and different classes

کلاس‌های ترکیبی و کلاس‌های مختلف

این بخش به تشخیص بیان‌های ترکیبی می‌پردازد، جایی که احساسات ترکیبی از بیان‌های اساسی هستند. نویسندگان روش خود را در دو سناریو مقایسه می‌کنند:

کلاس‌های ترکیبی: هنگام استفاده از بیان‌های ترکیبی RAF-DB به‌عنوان کلاس‌های فضای باز، روش نویسندگان AUROC برابر 0.771 را به‌دست می‌آورد که به‌طور قابل توجهی بهتر از سایر روش‌هاست.

کلاس‌های مختلف (AffectNet): نویسندگان همچنین مدل را بر روی AffectNet با استفاده از کلاس تحقیر به‌عنوان فضای باز آزمایش می‌کنند. در حالی که عملکرد تشخیص به دلیل برچسب‌های نویزی در AffectNet کاهش می‌یابد، روش آنها همچنان AUROC برابر 0.674 را به‌دست می‌آورد که بهترین میان روش‌های مقایسه‌شده است.

online application for on given sample

کاربرد آنلاین برای یک نمونه خاص

این بخش به توانایی روش در مدیریت تشخیص آنلاین می‌پردازد، جایی که هر نمونه به‌صورت جداگانه آزمایش می‌شود. آزمایش نشان می‌دهد:

روش تقریباً به خوبی در حالت آفلاین عمل می‌کند. AUROC در تشخیص آنلاین تنها 2.6٪ کاهش می‌یابد (از 0.909 به 0.883)، که همچنان بهتر از سایر روش‌های پیشرفته است.

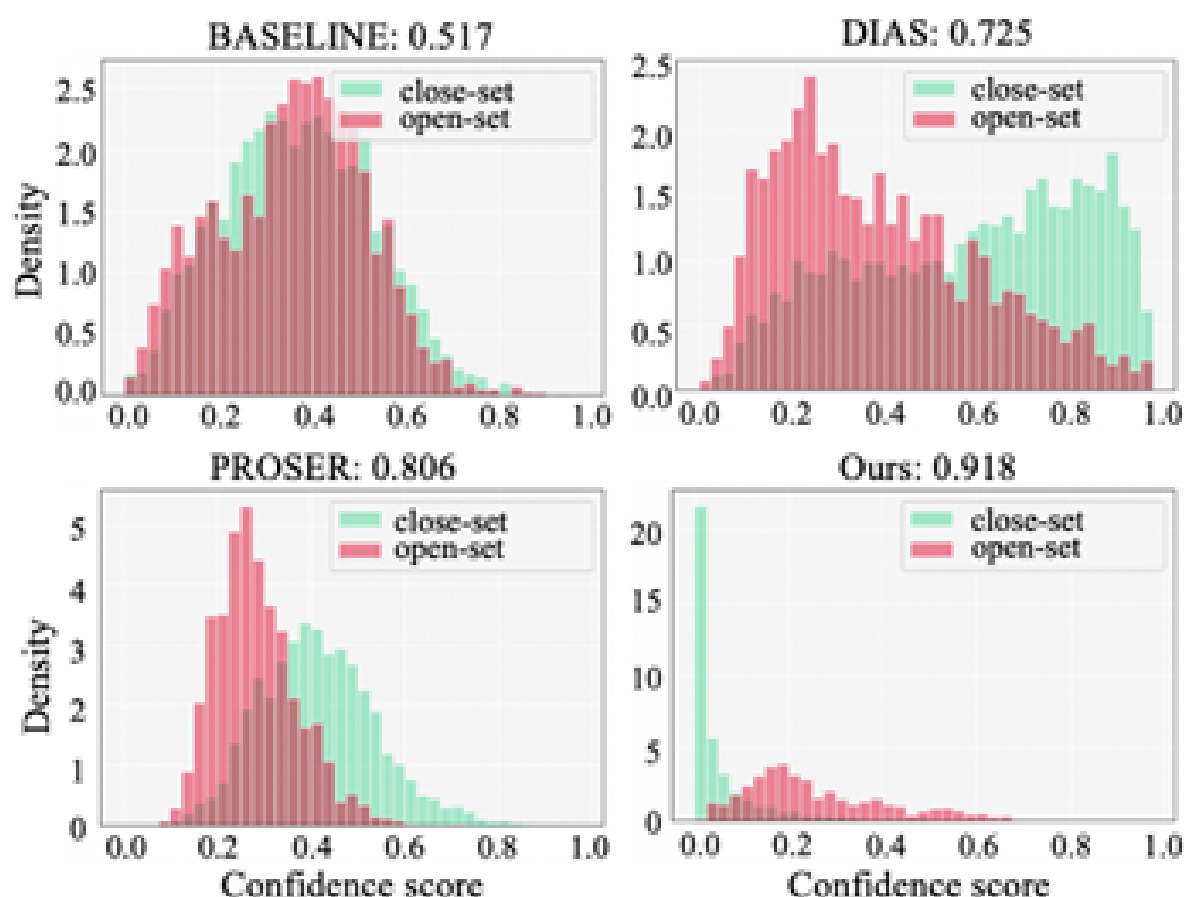
این موضوع کارآمدی مدل را حتی برای کاربردهای آنلاین و بلادرنگ نشان می‌دهد

Open class	AUROC	Open class	AUROC
Surprise	0.918/0.894	Sadness	0.911/0.882
Fear	0.907/0.870	Anger	0.906/0.898
Disgust	0.910/0.899	Neutral	0.917/0.889
Happiness	0.892/0.846	Mean	0.909/0.883

این جدول عملکرد روش پیشنهادی را در سناریوهای تشخیص آفلاین و آنلاین مقایسه می‌کند.

تشخیص آفلاین (جایی که مدل دوباره آموزش داده می‌شود) به AUROC برابر با ۰.۹۰۹ دست می‌یابد.

تشخیص آنلاین (جایی که مدل نیازی به آموزش مجدد ندارد) با کاهش جزئی مواجه می‌شود، به طوری که AUROC به ۰.۸۸۳ کاهش می‌یابد، اما همچنان از سایر روش‌ها عملکرد بهتری دارد.



تحلیل هایپرمترها (دو تصویر در سمت چپ)

این تصاویر نشان می‌دهند که تغییر وزن همسانی و تعداد دوره‌های آموزشی چگونه بر عملکرد AUROC برای شبکه‌های ResNet-18 و ResNet-50 تأثیر می‌گذارد.

وزن همسانی: افزایش وزن همسانی از ۱ به ۵ عملکرد را بهبود می‌بخشد، اما وقتی وزن بیش از حد بالا می‌رود، عملکرد اندکی کاهش می‌یابد.

دوره‌های آموزشی: عملکرد با افزایش تعداد دوره‌ها بهبود می‌یابد، اما در حدود ۴۰ دوره به حالت اشباع می‌رسد.

further analyses

تحلیل‌های بیشتر

این بخش شامل چندین زیر بخش است که به بینش‌های عمیق‌تر در مورد عملکرد روش می‌پردازد:

مطالعه ابلیشن (AUROC)

این جدول با ارزیابی مطالعات ابلیشن، به بررسی نقش مؤلفه‌های مختلف در روش پیشنهادی می‌پردازد:

بدون ماژول همسانی نقشه توجه، عملکرد به ۰.۵۱۷ کاهش می‌یابد.

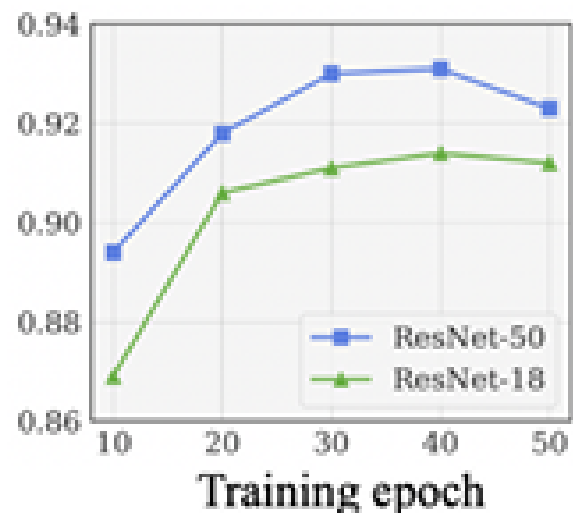
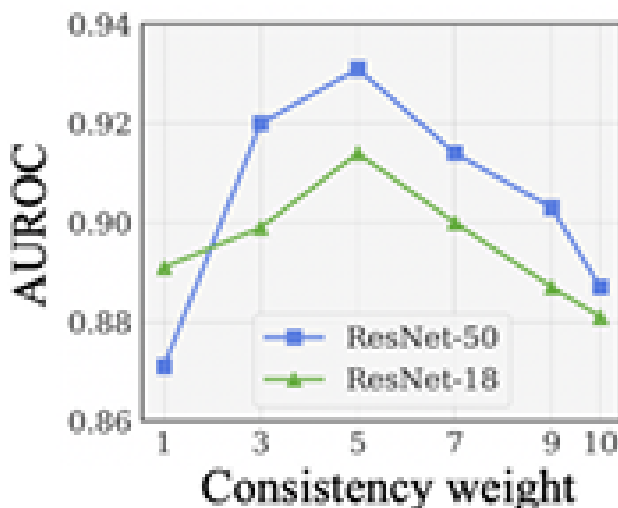
افزودن ماژول همسانی نقشه توجه عملکرد را به ۰.۸۸۵ افزایش می‌دهد.

ترکیب همسانی نقشه توجه با نرخ یادگیری چرخه‌ای و آموزش چرخه‌ای بهترین نتیجه را می‌دهد (AUROC = 0.918).

این جدول تأیید می‌کند که هر ماژول به عملکرد کلی کمک می‌کند.

Attention	Cycle LR	Cycle Training	AUROC
			0.517
✓			0.885
✓	✓		0.912
✓	✓	✓	0.918

Table 5: The ablation study of our method.



توزیع برچسب‌های شبه (دو تصویر در سمت راست)

این تصاویر توزیع برچسب‌های شبه را نمایش می‌دهند:

تصویر سمت چپ نشان می‌دهد که برچسب‌های شبه نمونه‌های باز-مجموعه به تمام کلاس‌های بسته-مجموعه توزیع می‌شوند.

تصویر سمت راست نشان می‌دهد که عملکرد به طور قابل توجهی کاهش می‌یابد وقتی برچسب‌های شبه در کلاس‌های معنایی مشابه متمرکز می‌شوند (به عنوان مثال، برچسب‌گذاری تمام نمونه‌های Surprise به عنوان Happiness).

visualization of confidence scores

بصری‌سازی نمرات اطمینان

نویسندگان نمرات اطمینان را با استفاده از یک نمودار بصری‌سازی می‌کنند که جداسازی نمونه‌های فضای باز و بسته را نشان می‌دهد. نکات کلیدی:

روش‌های پایه (مانند softmax) شکست می‌خورند زیرا نمرات اطمینان نمونه‌های فضای باز با نمونه‌های بسته همپوشانی دارند.

روش نویسندگان این همپوشانی را به‌طور قابل توجهی کاهش می‌دهد، با تبدیل مسئله Open-Set FER به مسئله تشخیص برچسب‌های نویزی، و این باعث بهبود تشخیص احساسات نادیده می‌شود.

ablation and hyperparameter study

مطالعه ابلیشن و فراسنچ‌ها

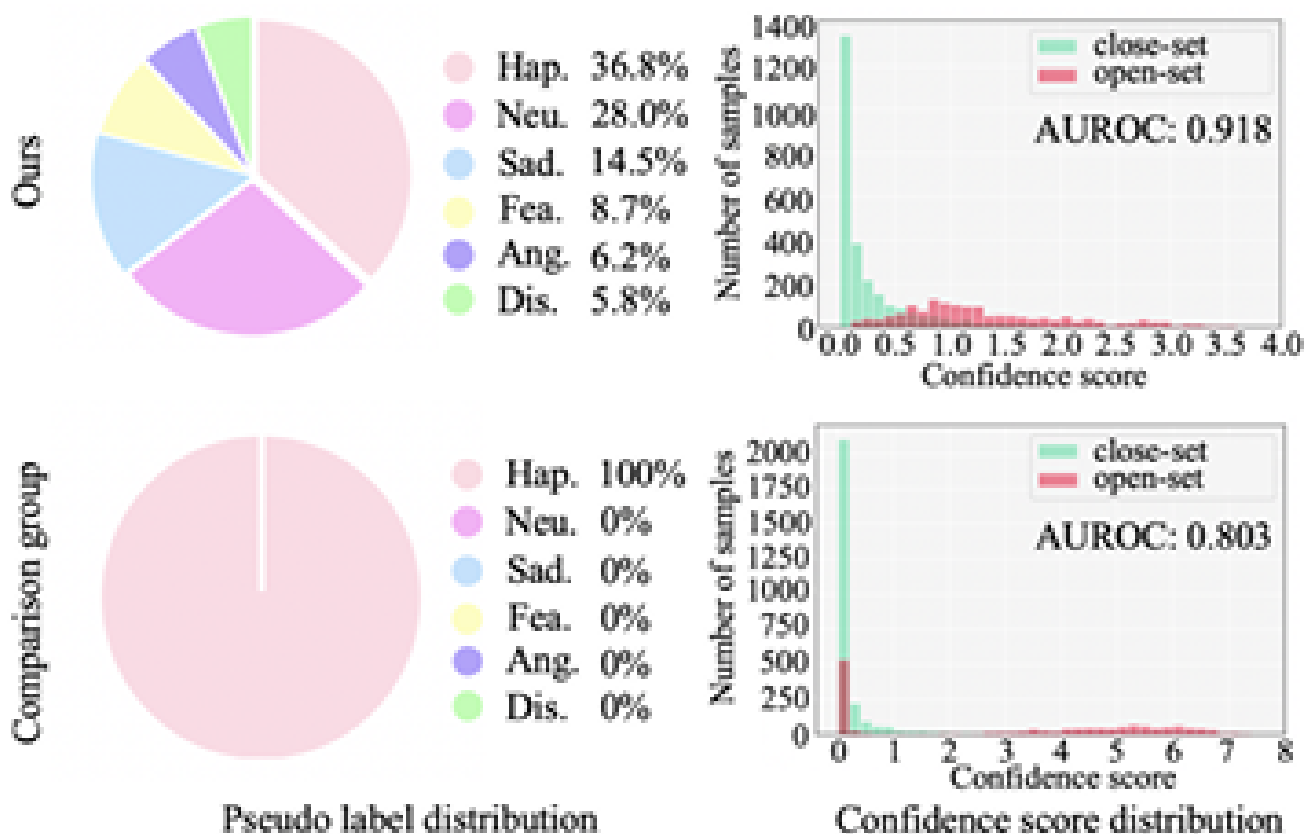
در این بخش، نویسندگان تأثیر ماژول‌های مختلف در روش خود را مطالعه می‌کنند. یافته‌های کلیدی:

سازگاری نقشه توجه مؤثرترین ماژول است که AUROC را از 0.517 به 0.885 بهبود می‌دهد.

آموزش چرخه‌ای AUROC را با تکرار بین دو مدل بیشتر بهبود می‌دهد (AUROC = 0.912).

ترکیب همه تکنیک‌ها (سازگاری نقشه توجه، نرخ یادگیری چرخه‌ای، آموزش چرخه‌ای) بهترین نتیجه را به‌دست می‌آورد (AUROC = 0.918).

نویسندگان همچنین مطالعه‌ای در مورد فراسنچ‌ها انجام داده‌اند که نشان می‌دهد روش آنها نسبت به تغییرات وزن سازگاری و دوره‌های آموزشی حساس نیست. ResNet-50 بهتر از ResNet-18 عمل می‌کند، اما نتایج ResNet-18 برای مقایسه منصفانه با سایر روش‌ها گزارش شده است.



. بصری سازی ویژگی های آموخته شده (دو تصویر در سمت چپ)

این تصاویر با استفاده از t-SNE ویژگی های آموخته شده از مدل پایه و روش پیشنهادی را به تصویر می کشند.

روش پایه ویژگی های باز-مجموعه و بسته-مجموعه را با هم ترکیب می کند و منجر به جداسازی ضعیف می شود.

روش پیشنهادی به طور مؤثر ویژگی های باز-مجموعه (با علامت قرمز) را از ویژگی های بسته-مجموعه جدا می کند، که نشان دهنده عملکرد بهتر است.

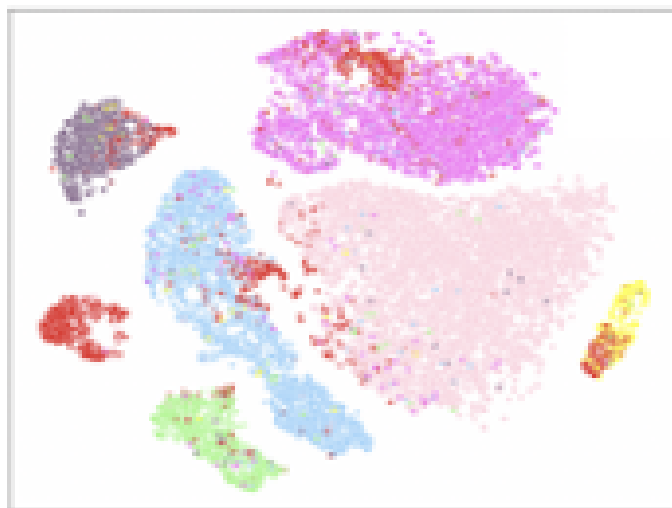
analyses of pseudo label distributions

حلیل توزیع برچسب های شبه

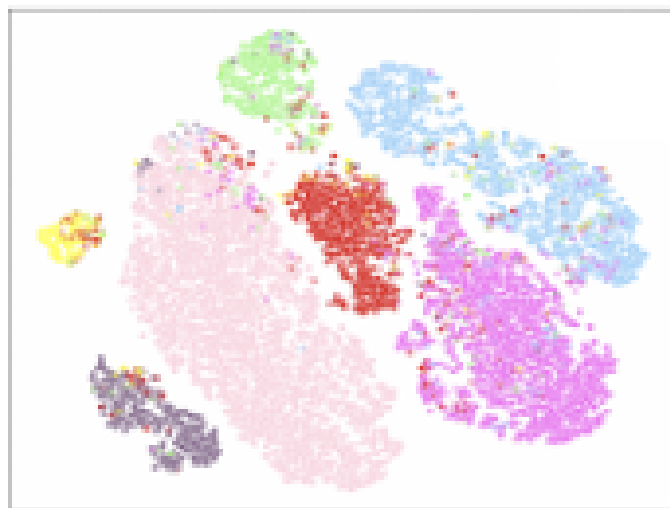
این زیر بخش به تحلیل توزیع برچسب های شبه می پردازد. در Open-Set FER، برچسب های شبه نمونه های فضای باز در بین همه کلاس های بسته توزیع می شوند. یافته های کلیدی:

این روش خوب عمل می کند زیرا برچسب های شبه به صورت متقارن در بین همه کلاس ها توزیع می شوند که تشخیص نمونه های فضای باز را آسان تر می کند.

در مقابل، اگر برچسب های شبه در یک کلاس متمرکز شوند (مثلاً همه نمونه های تعجب به عنوان خوشحالی طبقه بندی شوند)، عملکرد به طور قابل توجهی کاهش می یابد (AUROC از 0.918 به 0.803).



Baseline



Ours

بصری سازی امتیازات اطمینان (دو تصویر در سمت راست)

این تصاویر امتیازات اطمینان روش های مختلف شناسایی باز-مجموعه را نشان می دهند:

روش پایه همپوشانی قابل توجهی بین نمونه های بسته-مجموعه و باز-مجموعه دارد، که تشخیص بین آن ها را دشوار می سازد.

روش های DIAS و PROSER این همپوشانی را کاهش می دهند اما همچنان در جداسازی کامل نمونه ها ناموفق هستند.

روش پیشنهادی شناسایی حالات چهره باز-مجموعه را به یک مسئله تشخیص برچسب های نویزی تبدیل می کند و به جداسازی واضح تری بین داده های بسته-مجموعه و باز-مجموعه دست می یابد.

visualization of learned features

بصری سازی ویژگی های یادگرفته شده

نویسندگان از t-SNE برای بصری سازی ویژگی های یادگرفته شده مدل استفاده می کنند. مشاهدات کلیدی:

روش پایه در جدا کردن ویژگی های فضای باز از ویژگی های بسته شکست می خورد و آنها را با هم مخلوط می کند.

روش پیشنهادی به طور مؤثری ویژگی های فضای باز را از ویژگی های بسته جدا می کند، که نشان دهنده قدرت رویکرد آنها در جلوگیری از بیش برآزش به برچسب های اشتباه است.

conclusion

در نتیجه گیری، نویسندگان مشارکت های خود را خلاصه می کنند:

آنها یک کار جدید را پیشنهاد می کنند: تشخیص بیان چهره فضای باز (FER).

آنها شناسایی می‌کنند که فاصله کم بین کلاس‌ها در داده‌های بیان چهره باعث کاهش عملکرد روش‌های موجود تشخیص فضای باز می‌شود.

با تبدیل Open-Set FER به مسئله تشخیص برچسب‌های نویزی، روش جدید آنها از برچسب‌های شبه، سازگاری نقشه‌های توجه و آموزش چرخه‌ای برای بهبود تشخیص فضای باز استفاده می‌کند.

آزمایش‌های گسترده نشان می‌دهد که روش آنها به‌طور قابل توجهی بهتر از روش‌های پیشرفته است و حتی در سناریوهای چالش‌برانگیز مانند بیان‌های ترکیبی و تشخیص آنلاین خوب عمل می‌کند.

EXPLANATIONS

Open Set یا "مجموعه باز" در یادگیری ماشین و تشخیص الگوها به شرایطی اشاره دارد که مدل تنها با بخشی از تمام دسته‌های ممکن در طول آموزش مواجه است، اما در مرحله تست با داده‌هایی از دسته‌های ناشناخته برخورد می‌کند. به عبارت دیگر، در یک سیستم Open Set، مدل باید قادر باشد نه تنها داده‌هایی را که در دسته‌های آموزش دیده هستند، شناسایی کند، بلکه بتواند موارد ناشناخته یا غیرمربوط به کلاس‌های شناخته شده را نیز شناسایی کند.

توضیح مفصل:

در اکثر مسائل یادگیری ماشین سنتی (مانند طبقه‌بندی)، فرض بر این است که همه‌ی دسته‌ها یا کلاس‌ها که مدل ممکن است با آنها مواجه شود، از پیش در داده‌های آموزشی مشخص شده‌اند. این وضعیت به عنوان Closed Set یا "مجموعه بسته" شناخته می‌شود، جایی که دسته‌ها در مرحله آموزش و تست یکی هستند. اما در دنیای واقعی، سیستم‌ها اغلب با داده‌هایی برخورد می‌کنند که به دسته‌هایی تعلق دارند که در طول آموزش دیده نشده‌اند. در چنین شرایطی، مدل باید بتواند به درستی تشخیص دهد که داده ورودی به دسته‌ای ناشناخته تعلق دارد.

close-set #

کلاس‌هایی که مدل باید در مرحله تست یا پیش‌بینی با آنها

کار کند، از قبل در مرحله آموزش دیده شده‌اند. به بیان ساده‌تر، مدل یادگیری ماشین در مواجهه با داده‌های تست فقط با کلاس‌ها و برچسب‌هایی سر و کار دارد که از قبل در داده‌های آموزشی وجود داشته‌اند و هیچ کلاس ناشناخته‌ای در این مجموعه وجود ندارد.

ویژگی‌های اصلی Close-Set:

کلاس‌های مشخص و محدود: در یک مسئله مجموعه بسته، مدل تنها با تعدادی کلاس از پیش تعریف شده در طول آموزش و تست سروکار دارد. هیچ داده یا دسته جدیدی که مدل آن را ندیده باشد، وجود نخواهد داشت.

آموزش و تست از کلاس‌های مشترک: در یک سناریوی Close-Set، داده‌های آموزشی و تست هر دو از همان کلاس‌ها تشکیل شده‌اند. برای مثال، اگر مدل بر روی دسته‌های "گره"، "سگ"، و "اسب" آموزش دیده باشد، در مرحله تست نیز تنها

داده‌هایی مربوط به همین دسته‌ها را خواهد دید.

چالش کمتر نسبت به Open-Set: تشخیص در مجموعه بسته معمولاً ساده‌تر است، زیرا مدل فقط باید داده‌های دیده‌شده را طبقه‌بندی کند و نیازی به شناسایی داده‌های ناشناخته ندارد.

noisy detection

مسئله تشخیص برچسب‌های نویزی (Noisy Label Detection) به مشکلی در یادگیری ماشین اشاره دارد که در آن داده‌های آموزشی حاوی برچسب‌های اشتباه یا نادرست هستند. در یادگیری نظارت‌شده، مدل‌ها از داده‌های برچسب‌گذاری‌شده برای یادگیری الگوها استفاده می‌کنند، اما وقتی برچسب‌ها حاوی نویز باشند (یعنی برچسب اشتباه یا غلط به داده تخصیص داده شده باشد)، عملکرد مدل به شدت تحت تأثیر قرار می‌گیرد و دقت آن کاهش می‌یابد.

منابع برچسب‌های نویزی:

اشتباه انسانی: در فرآیند برچسب‌گذاری داده‌ها، ممکن است افرادی که داده‌ها را برچسب‌گذاری می‌کنند اشتباه کنند. این موضوع در مجموعه داده‌های بزرگ و پیچیده شایع است.

خودکارسازی ناقص: گاهی از روش‌های خودکار برای برچسب‌گذاری داده‌ها استفاده می‌شود، که ممکن است خطاهایی در آن وجود داشته باشد.

ابهام در داده‌ها: برخی داده‌ها ممکن است به راحتی قابل دسته‌بندی نباشند یا خود داده مبهم باشد، مانند تصاویر یا متون مبهم که نمی‌توان با اطمینان برچسب صحیحی به آن‌ها داد.

اثرات برچسب‌های نویزی:

کاهش دقت مدل: برچسب‌های نویزی باعث می‌شوند مدل الگوهای اشتباهی را یاد بگیرد و نتواند به درستی تعمیم دهد. در نهایت، این موضوع باعث کاهش دقت مدل در پیش‌بینی‌ها می‌شود.

افزایش زمان و هزینه آموزش: زمانی که داده‌های نویزی به مدل داده می‌شود، زمان آموزش بیشتر می‌شود و مدل ممکن است به پارامترهای نامناسب تنظیم شود که نیاز به تصحیح دارد.

عدم تعمیم‌پذیری مناسب: مدل‌هایی که با برچسب‌های نویزی آموزش می‌بینند، توانایی تعمیم‌دهی به داده‌های جدید و ناشناخته را از دست می‌دهند.

attention mechanism

مکانیزم توجه (Attention Mechanism) یک تکنیک پیشرفته در یادگیری ماشین و به‌خصوص در مدل‌های عصبی است که به مدل کمک می‌کند تا بتواند روی بخش‌های مهم‌تری از ورودی تمرکز کند و اهمیت آن‌ها را درک نماید. این مکانیزم به‌طور گسترده در مسائلی مانند ترجمه ماشینی، پردازش زبان طبیعی (NLP)، و بینایی ماشین مورد استفاده قرار می‌گیرد.

چرا مکانیزم توجه به وجود آمد؟

در مسائل پیچیده‌ای مانند پردازش زبان طبیعی یا بینایی ماشین، هر ورودی دارای حجم زیادی از اطلاعات است که همگی دارای اهمیت یکسانی نیستند. برای مثال، در یک جمله، ممکن است برخی از کلمات مهم‌تر باشند و مدل نیاز داشته باشد تا توجه بیشتری به آن‌ها بکند. مکانیزم توجه به مدل اجازه می‌دهد تا به جای در نظر گرفتن تمام ورودی‌ها به صورت یکسان، روی قسمت‌هایی که برای وظیفه فعلی مهم‌تر هستند تمرکز کند و آن‌ها را اولویت‌بندی نماید.

مفهوم کلی مکانیزم توجه

مکانیزم توجه به هر قسمت از داده‌های ورودی، یک وزن اختصاص می‌دهد که این وزن نشان‌دهنده میزان اهمیت آن قسمت است. سپس مدل بیشتر روی قسمت‌هایی که وزن بالاتری دارند تمرکز می‌کند و از آن‌ها برای تصمیم‌گیری استفاده می‌کند.

Attention Map Consistency به معنای ****یکسانی یا سازگاری نقشه توجه**** در زمینه‌های مختلف یادگیری ماشین، به ویژه در مدل‌های مبتنی بر مکانیزم توجه (Attention Mechanism)

(مانند شبکه‌های عصبی بازگشتی (RNN)، شبکه‌های عصبی کانولوشنی (CNN)، و ترانسفورمرها (Transformers)) است.

مکانیزم توجه به مدل‌ها کمک می‌کند تا بتوانند روی قسمت‌های مهم داده‌های ورودی تمرکز کنند و ارتباطات یا ویژگی‌های کلیدی را بهتر درک کنند. ****نقشه توجه (Attention Map)**** نشان می‌دهد که مدل در هر گام از پردازش، چقدر توجه خود را به بخش‌های مختلف داده‌ها معطوف کرده است.

تعریف Attention Map Consistency

Attention Map Consistency به این موضوع اشاره دارد که مدل در طی پردازش ورودی‌های مشابه یا تکراری، بتواند الگوی توجه یکسان یا سازگاری را حفظ کند. این سازگاری به معنای آن است که مدل در پردازش‌های مختلف روی همان داده، همان بخش‌های مهم یا ویژگی‌های کلیدی را شناسایی کرده و به آن‌ها توجه می‌کند.

اهمیت Attention Map Consistency

- ****پایداری و اعتماد به مدل****: اگر یک مدل یادگیری ماشین بتواند نقشه توجه پایدار و سازگاری را برای ورودی‌های مشابه ارائه دهد، نشان‌دهنده این است که مدل به درستی یاد گرفته است کدام قسمت‌های داده مهم هستند. این موضوع برای کاربردهایی مانند بینایی ماشین و پردازش زبان طبیعی که نیاز به تمرکز بر جزئیات خاص دارند، اهمیت زیادی دارد.

- ****تعمیم‌پذیری بهتر****: سازگاری در نقشه توجه می‌تواند نشان‌دهنده این باشد که مدل توانایی تعمیم‌پذیری به داده‌های مشابه یا داده‌های جدید را دارد و می‌تواند در شرایط مختلف عملکرد ثابتی از خود نشان دهد.

cycle training

Cycle Training یا آموزش چرخه‌ای یک روش یا تکنیک در آموزش مدل‌های یادگیری ماشین است که طی آن، داده‌های آموزشی به صورت دوره‌ای یا چرخه‌ای به مدل ارائه می‌شوند تا مدل بتواند با به‌روزرسانی‌های پیوسته و تکراری، بهبود یابد. این روش به‌ویژه در یادگیری نیمه‌نظارتی (Semi-supervised Learning)، یادگیری بدون نظارت (Unsupervised Learning)، و در برخی معماری‌های خاص مانند شبکه‌های مولد تقابلی (GANs) استفاده می‌شود.

اصول Cycle Training:

آموزش چرخه‌ای شامل تکرار مداوم یک فرآیند آموزشی است که می‌تواند شامل تغییرات در داده‌ها یا مدل باشد. هدف این است که با هر چرخه جدید، مدل بهبود یابد و توانایی خود را برای تعمیم به داده‌های جدید افزایش دهد.

کاربردهای Cycle Training:

1. CycleGAN: یکی از معروف‌ترین کاربردهای آموزش چرخه‌ای در CycleGAN است، یک نوع شبکه مولد تقابلی (Generative Adversarial Network) که برای تبدیل تصویر به تصویر بدون استفاده از برچسب‌ها استفاده می‌شود. در CycleGAN، چرخه‌های متعددی از آموزش وجود دارد که به مدل کمک می‌کند تا بتواند یاد بگیرد چگونه تصاویر را از یک دامنه به دامنه دیگر تبدیل کند و در عین حال ثبات و سازگاری آن‌ها را حفظ کند. به عنوان مثال، تبدیل تصویر از دامنه «اسب» به «گورخر» و بالعکس.

مکانیزم CycleGAN دو مولد (Generator) و دو متمایزکننده (Discriminator) دارد که در چرخه‌های جداگانه آموزش می‌بینند:

o مولد اول وظیفه تبدیل یک تصویر از دامنه A به دامنه B را دارد.

o مولد دوم تصویر تولید شده را دوباره به دامنه A بازمی‌گرداند.

o متمایزکننده‌ها بررسی می‌کنند که آیا تصاویر تولیدشده واقعی هستند یا نه، و به بهبود عملکرد مولدها کمک می‌کنند. این چرخه‌ها به مدل اجازه می‌دهند که یاد بگیرد چگونه دو دامنه کاملاً متفاوت را به یکدیگر تبدیل کند، در حالی که ویژگی‌های اصلی هر دامنه حفظ می‌شوند.

2. آموزش با داده‌های مصنوعی: در برخی موارد، از آموزش چرخه‌ای برای ترکیب داده‌های مصنوعی و واقعی استفاده می‌شود. به این صورت که در هر چرخه جدید، داده‌های مصنوعی بیشتری تولید و به مدل اضافه می‌شود تا مدل به صورت تدریجی و مداوم بهتر شود.

3. یادگیری چند وظیفه‌ای (Multi-task Learning): در یادگیری چند وظیفه‌ای، مدل ممکن است در چرخه‌های مختلف برای وظایف مختلف آموزش ببیند. این چرخه‌ها به مدل کمک می‌کنند تا از داده‌های مربوط به هر وظیفه یاد بگیرد و در عین حال ارتباط بین وظایف مختلف را حفظ کند.

مزایای Cycle Training:

1. بهبود تدریجی مدل: با هر چرخه جدید، مدل از داده‌ها و خطاهای چرخه قبلی یاد می‌گیرد و عملکرد خود را بهبود می‌دهد.

2. حفظ تعادل بین دامنه‌ها: در مدل‌های مانند CycleGAN، چرخه‌های آموزشی کمک می‌کنند که مدل تعادل بین دو دامنه مختلف (مانند تبدیل تصویر از یک سبک به سبک دیگر) را حفظ کند.

3. استفاده از داده‌های بدون برچسب: در برخی از روش‌های نیمه‌نظارتی یا بدون نظارت، آموزش چرخه‌ای به مدل اجازه می‌دهد که از داده‌های بدون برچسب یا داده‌های تولید شده توسط خود مدل، به طور موثر یاد بگیرد.

محدودیت‌ها:

• پیچیدگی محاسباتی: آموزش چرخه‌ای می‌تواند از لحاظ محاسباتی پیچیده و زمان‌بر باشد، به‌ویژه در مدل‌های بزرگی مانند

CycleGAN که نیاز به چندین مولد و متمایزکننده دارد.

• تنظیم پارامترها: برای دستیابی به عملکرد بهینه، نیاز به تنظیم دقیق پارامترها و کنترل کردن روند چرخه‌های آموزشی وجود دارد.

نتیجه‌گیری:

Cycle Training یا آموزش چرخه‌ای یک تکنیک مفید در یادگیری ماشین است که به مدل‌ها اجازه می‌دهد از تکرارهای مداوم و بازخوردهای چرخه‌ای برای بهبود عملکرد خود استفاده کنند. این روش به‌ویژه در مسائل بدون نظارت و نیمه‌نظارتی مانند CycleGAN کاربرد دارد و می‌تواند منجر به تعمیم‌پذیری بهتر مدل در شرایط مختلف شود.

در ادامه هر یک از این اصطلاحات با جزئیات بیشتر و عمق بیشتری توضیح داده می‌شوند:

1. **OSR (Open-Set Recognition - تشخیص مجموعه باز)**

تشخیص مجموعه باز یک مسئله پیچیده در یادگیری ماشین است که در آن سیستم باید توانایی تمایز بین کلاس‌های شناخته‌شده (کلاس‌هایی که در مرحله آموزش مشاهده کرده) و کلاس‌های ناشناخته (کلاس‌هایی که در مرحله آموزش ندیده) را داشته باشد. این مسئله در بسیاری از کاربردهای واقعی مانند تشخیص هویت، تشخیص چهره، و امنیت بسیار مهم است زیرا در این حوزه‌ها ممکن است نمونه‌هایی وجود داشته باشند که در مرحله آموزش دیده نشده‌اند. مدل‌های OSR باید قادر باشند کلاس‌های جدید را تشخیص داده و آن‌ها را به عنوان ناشناخته دسته‌بندی کنند، به جای این که به اشتباه آن‌ها را به یکی از کلاس‌های شناخته‌شده اختصاص دهند.

2. **Generative Models (مدل‌های مولد)**

مدل‌های مولد نوعی از مدل‌های یادگیری ماشین هستند که به جای تمرکز بر طبقه‌بندی یا پیش‌بینی، به تولید داده‌های جدید بر اساس توزیع داده‌های موجود می‌پردازند. این مدل‌ها داده‌هایی شبیه به داده‌های اصلی تولید می‌کنند. به عنوان مثال، در مدل‌های مولد تصاویر، این مدل‌ها می‌توانند تصاویر جدیدی تولید کنند که شباهت زیادی به تصاویر واقعی داشته باشند. کاربرد این مدل‌ها شامل تولید تصاویر مصنوعی، افزایش داده (data augmentation)، و حتی در زمینه‌هایی مثل خلق آثار هنری و تصاویر خلاقانه است.

3. **GANs (Generative Adversarial Networks - شبکه‌های مولد متخاصم)**

شبکه‌های مولد متخاصم یکی از موفق‌ترین مدل‌های مولد هستند. در GAN، دو شبکه عصبی با هم رقابت می‌کنند:

- **Generator (مولد)**: سعی می‌کند داده‌های مصنوعی تولید کند که شبیه به داده‌های واقعی باشند.

- **Discriminator (تمایزگر)**: تلاش می‌کند داده‌های مصنوعی تولیدشده توسط مولد را از داده‌های واقعی تمایز دهد.

این رقابت موجب بهبود مستمر هر دو شبکه می‌شود، و نتیجه آن یک مولد قدرتمند است که می‌تواند داده‌هایی تولید کند که بسیار شبیه به داده‌های واقعی باشند. این مدل‌ها در زمینه‌هایی مانند تولید تصاویر چهره، طراحی محصولات، و حتی تولید داده‌های پزشکی کاربرد دارند.

4. Inter-class Distance (فاصله بین کلاسی)**

فاصله بین کلاسی به فاصله بین بردارهای ویژگی‌هایی که به کلاس‌های مختلف تعلق دارند اشاره دارد. این مفهوم اهمیت زیادی در یادگیری ماشین دارد، زیرا اگر فاصله بین کلاسی بزرگ باشد، مدل می‌تواند به راحتی بین کلاس‌های مختلف تمایز قائل شود و دقت بالاتری خواهد داشت. از طرف دیگر، اگر فاصله بین کلاسی کوچک باشد، مدل ممکن است نتواند به خوبی بین کلاس‌ها تفکیک کند و دچار خطاهای طبقه‌بندی شود. هدف بسیاری از الگوریتم‌های یادگیری، افزایش این فاصله برای بهبود عملکرد طبقه‌بندی‌کننده‌ها است.

5. Classifier Overconfidence (اعتماد بیش از حد طبقه‌بندی‌کننده)**

این اصطلاح به وضعیتی اشاره دارد که مدل طبقه‌بندی‌کننده به طور غیرمنطقی اعتماد زیادی به پیش‌بینی‌های خود دارد، حتی در مواردی که ممکن است داده‌های ورودی متعلق به یک کلاس ناشناخته یا اشتباه باشند. این مشکل معمولاً در مدل‌هایی که روی مجموعه‌های بسته آموزش دیده‌اند بیشتر دیده می‌شود. اعتماد بیش از حد می‌تواند منجر به پیش‌بینی‌های اشتباه و نادرست شود، مخصوصاً در مواردی که داده‌های ورودی دارای نویز هستند یا به دسته‌های ناشناخته تعلق دارند.

6. Attention Map Consistency (سازگاری نقشه توجه)**

نقشه توجه در مدل‌های یادگیری عمیق نشان می‌دهد که مدل به کدام بخش از داده‌ها یا تصاویر بیشتر توجه می‌کند. سازگاری نقشه توجه به این معنا است که مدل در مراحل مختلف آموزش یا در برخورد با داده‌های مختلف باید بتواند توجه خود را به بخش‌های مهم داده‌ها متمرکز کند و این تمرکز را به صورت پایدار و منسجم حفظ کند. این ویژگی در مدل‌های تشخیص حالات چهره اهمیت زیادی دارد، زیرا مدل باید همواره به بخش‌های مهم چهره (مانند چشم‌ها، دهان و ابروها) توجه کند تا بتواند حالات چهره را به درستی تشخیص دهد.

7. Cycle Learning Rate (نرخ یادگیری چرخه‌ای)**

نرخ یادگیری چرخه‌ای یک تکنیک برای تنظیم نرخ یادگیری در طول فرآیند آموزش است. به جای استفاده از یک نرخ یادگیری ثابت یا کاهش پیوسته آن، نرخ یادگیری در این روش به صورت چرخه‌ای تغییر می‌کند. این تغییرات چرخه‌ای می‌توانند به مدل کمک کنند که از افتادن در کمینه‌های محلی (local minima) جلوگیری کند و در عین حال بهبود عملکرد کلی را افزایش دهد. این روش در شبکه‌های عصبی عمیق بسیار مؤثر بوده و در مواردی مانند یادگیری انتقالی نیز استفاده می‌شود.

8. Ensemble Learning (یادگیری مجموعه‌ای)**

یادگیری مجموعه‌ای یک تکنیک در یادگیری ماشین است که به جای استفاده از یک مدل، چندین مدل را ترکیب می‌کند تا عملکرد کلی بهبود یابد. ایده اصلی پشت این روش این است که هر مدل ممکن است نقاط ضعف خاص خود را داشته باشد، اما ترکیب آن‌ها می‌تواند خطاها را کاهش دهد. روش‌های مختلفی برای یادگیری مجموعه‌ای وجود دارد:

- **Bagging**: مدل‌های مختلف روی زیرمجموعه‌های مختلفی از داده‌ها آموزش داده می‌شوند.

- **Boosting**: مدل‌ها به صورت زنجیره‌ای آموزش داده می‌شوند و هر مدل خطاهای مدل قبلی را بهبود می‌دهد.

این روش‌ها به‌ویژه در مسائل پیچیده و پرچالش مانند تشخیص حالات چهره کاربرد دارند.

9. ### **Classification Loss** (زیان طبقه‌بندی)**

زیان طبقه‌بندی معیار خطایی است که نشان می‌دهد مدل چقدر از پیش‌بینی درست دور است. انواع مختلفی از زیان طبقه‌بندی وجود دارد:

- **Cross-Entropy Loss**: رایج‌ترین تابع زیان برای مسائل طبقه‌بندی، که به مدل کمک می‌کند فاصله بین توزیع پیش‌بینی شده و توزیع واقعی را کاهش دهد.

- **Hinge Loss**: برای الگوریتم‌هایی مانند ماشین بردار پشتیبانی (SVM) استفاده می‌شود و بر اساس فاصله کلاس‌ها از مرز تصمیم‌گیری محاسبه می‌شود.

10. ### **High Loss** (زیان بالا) و **Low Loss** (زیان کم)**

- **High Loss** (زیان بالا): زمانی که مدل پیش‌بینی اشتباه می‌کند و اختلاف زیادی با برچسب واقعی دارد، زیان بالا محاسبه می‌شود.

- **Low Loss** (زیان کم): زمانی که مدل پیش‌بینی درستی دارد یا به پیش‌بینی درست نزدیک است، زیان کم خواهد بود.

11. ### **RAF-DB**

RAF-DB یک پایگاه داده بزرگ برای تشخیص حالات چهره است که شامل تصاویر چهره‌های مختلف از افراد در حالات احساسی مختلف است. این مجموعه داده برای تمرین و ارزیابی مدل‌های FER بسیار کاربرد دارد.

12. ### **FERPlus**

FERPlus نسخه گسترش‌یافته‌ای از مجموعه داده‌های قبلی FER است که شامل برچسب‌های احساسی دقیق‌تری است. این مجموعه داده شامل تصاویر با برچسب‌های احساسی است که دقت مدل‌های FER را بهبود می‌بخشد.

13. ### **AUROC (Area Under Receiver Operating Characteristic)**

AUROC یک معیار برای ارزیابی مدل‌های طبقه‌بندی دودویی است. این شاخص ناحیه زیر منحنی ROC را محاسبه می‌کند که نشان‌دهنده تعادل بین نرخ مثبت درست و نرخ منفی درست است. هرچه مقدار AUROC بالاتر باشد، مدل در تشخیص درست کلاس‌ها بهتر عمل می‌کند.

14. ### **FPR@TRR95**

این معیار نرخ مثبت کاذب (False Positive Rate) را در حالتی که نرخ مثبت درست (True Positive Rate) برابر با ۹۵٪ است اندازه‌گیری می‌کند. این معیار به ارزیابی دقت و عملکرد مدل در شرایط خاص کمک می‌کند.

15. ### **DIAS (Deep Image-to-Action Switching)**

DIAS روشی است که به مدل‌ها امکان می‌دهد از تحلیل عمیق تصاویر برای تشخیص و تغییر سریع در رفتار یا تصمیم‌گیری استفاده کنند. این روش به ویژه در سیستم‌های تعاملی یا بلادرنگ کاربرد دارد.

16. ### **EOW (End-of-Word)**

EOW به پایان کلمه در پردازش زبان طبیعی اشاره دارد. این مفهوم زمانی اهمیت دارد که سیستم‌ها باید پایان یک کلمه را تشخیص دهند تا پردازش مناسب روی جملات یا کلمات انجام شود.

17. ### **ProSER (Prototype-based Self-Reflective Learning)**

ProSER یک روش یادگیری مبتنی بر پروتوتایپ است که از نمونه‌های اصلی (پروتوتایپ‌ها) برای دسته‌بندی و تشخیص داده‌ها استفاده می‌کند. این روش کمک می‌کند تا مدل

ها بتوانند به صورت انعکاسی و با استفاده از نمونه‌های شاخص یاد بگیرند و عملکرد خود را بهبود دهند.

این توضیحات دقیق‌تر به شما کمک می‌کنند تا مفاهیم مختلف مرتبط با تشخیص حالات چهره و پروژه‌های FER را بهتر درک کنید و ارتباط آن‌ها با مسائل علمی و فنی را بفهمید.
