Weekly Report

|  |  |
| --- | --- |
| Project title | Realtime Face Recognition in Unconstraint Environments |
| Report period | Weekly |
| Name | Sajjad Aemmi |
| Date | 2021/05/19 |

# Follow-up on tasks from last week’s meeting

در گزارش قبلی مقاله های جدید در رابطه با معماری MobileNet را بررسی کردیم و مزایا و معایب هر روش را جداگانه دسته بندی کردیم.

نتیجه گیری کلی از مقالات فوق به شرح زیر است:

# روش های بررسی شده در بالا به طور کلی یکی از معایب زیر را دارند:

# الف) بر روی مجموعه داده های عمومی نظیر image net آموزش داده شده اند. و در مسئله چهره مورد استفاده قرار نگرفته اند.

# ب) با تغییر در معماری mobilenet سعی در کاهش تعداد پارامتر ها و افزایش سرعت را داشته اند. اما منجر به کاهش دقت شده اند.

# ج) برخی روش ها نظیر mobilenet v3 معایب بالا را ندارند. اما تعداد دسته ها در آنها برای آموزش ثابت می باشد. بر روی one shot learning استفاده نشده اند. و از داشتن لایه توجه بی بهره بوده اند. بنابرین در بعضی مجموعه داده های آزمایش معروف مانند Mega face به دقت بالای ۷۶ درصد نرسیده اند.

# سپس در مورد ایده جدید صحبت کردیم:

# استفاده از نسخه سوم معماری mobilenet به همراه افزودن ماژول Shuffle Attention که در سال ۲۰۲۱ ارائه شده است. می توان این مدل را به روش سایامیس و تابع ضرر arc face آموزش داد. استفاده از روش سایامیس کمک می کند تا تعداد دسته ها پس از آموزش قابل تغییر باشد و همچنین برای one shot learning بسیار مناسب است. و تابع ضرر arc face باعث افزایش فاصله بردار ویژگی های برون کلاسی و کاهش فاصله بردار ویژگی های درون کلاس میشود. این ایده به دلیل استفاده از لایه های توجه، می تواند همچنین برای چهره هایی که با ماسک پوشانده شده اند نیز، نسبت به روش های موجود دقت را افزایش دهد.

# همچنین اگر به صورت همزمان از سه مجموعه داده vgg face , iran celeb , Asian face برای آموزش استفاده کنیم. می توانیم انتظار دقت بیشتری از شبکه داشته باشیم. زیرا Asian face شامل چهره های آسیای شرقی مثل چین و ژاپن می باشد. و ... بدین ترتیب شبکه بر روی نژاد خاصی بایاس نخواهد شد. و افزایش دقت را به همراه خواهد داشت.

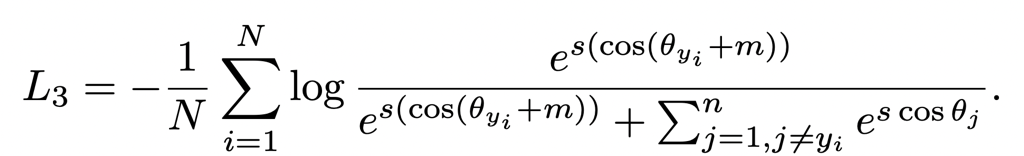
مقاله های اصلی که بناست در این کار مورد استفاده قرار بگیرند در پایین آورده شده اند.

برای تابع ضرر

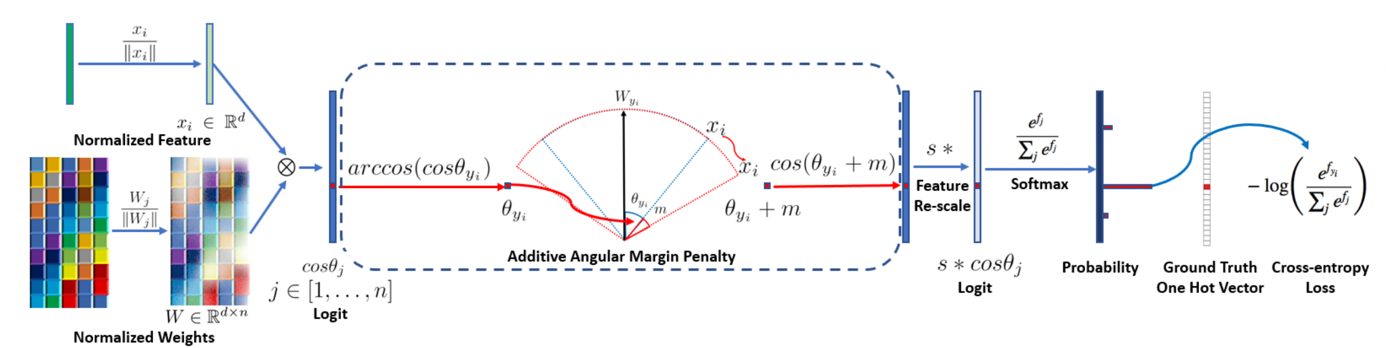
1. ArcFace: Additive Angular Margin Loss for Deep Face Recognition – 2019

url: <https://arxiv.org/pdf/1801.07698.pdf>

این روش که به دنبال مقاله cosFace آمده است. سعی در بهبود روش پیشین و افزایش دقت کرده است. بر همین اساس با طراحی تابع ضرر زیر با نام arcFace توانسته دقت بالایی بر روی دیتاست های مشهور کسب نماید:

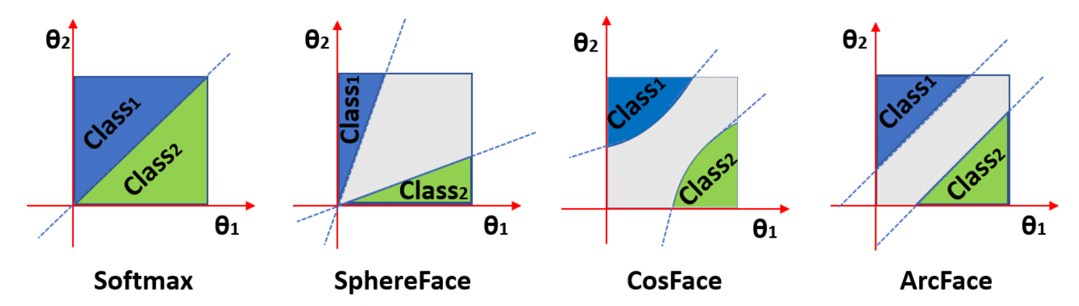


خلاصه این روش در شکل زیر آمده است:



مزایای این روش:

مقایسه دسته بندی این روش و روش های قبلی:



دقت 99.53 بر روی LFW با معماری ResNet 50

معایب این روش:

تعداد دسته ها (number of classes) ثابت است و پس از آموزش قابل تغییر نیست.

عدم سرعت بالای الگوریتم و توجه به realtime

برای ماژول Shuffle Attention

# SA-Net: Shuffle Attention for Deep Convolutional Neural Networks - 2021

url: <https://arxiv.org/abs/2102.00240>

لایه های توجه attention layers، شبکه عصبی را قادر می سازد به طور دقیق بر روی بخش های با اهمیت تر داده ورودی متمرکز شود. عمدتا دو مکانیسم توجه به طور گسترده در بینایی ماشین مورد استفاده قرار می گیرد

* spatial attention
* channel attention

که هدف آنها به ترتیب یافتن وابستگی همسایگی و وابستگی کانال در سطح پیکسل است. تلفیق آنها ممکن است عملکرد بهتری نسبت به پیاده سازی های منفرد آنها به دست آورد ، اما سربار محاسباتی را افزایش می دهد.

در این مقاله، یک ماژول Shuffle attention (SA) برای این مسئله پیشنهاد شده است. که برای ترکیب دو نوع attention layer با بار محاسباتی کم طراحی شده ست. SA ابتدا عمق کانال را به چندین ویژگی فرعی قبل از پردازش موازی آنها تقسیم می کند. سپس ، برای هر زیر ویژگی از یک واحد Shuffle برای وابستگی های هر دو بعد مکانی و کانال استفاده می کند. پس از آن ، همه زیر ویژگی ها جمع می شوند و برای امکان ادغام اطلاعاتی بین ویژگی های فرعی مختلف به کار گرفته می شود.

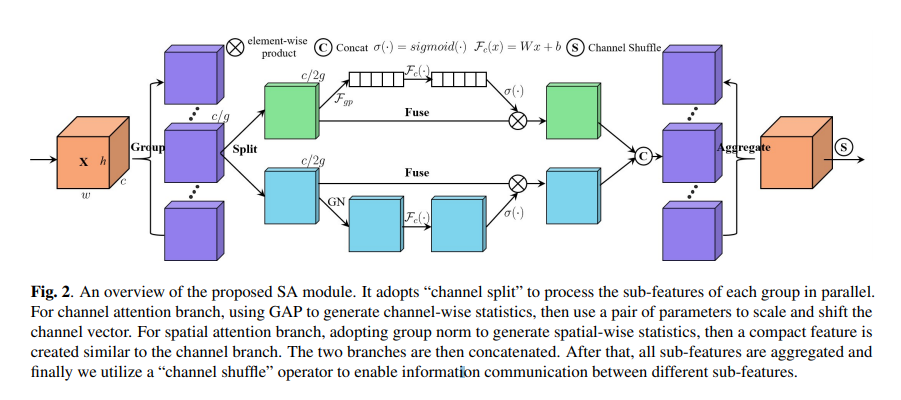
این روش بار محاسباتی بسیار کمی دارد. به عنوان مثال ، پارامترها و محاسبات SA در ResNet50 به ترتیب 300 در مقابل 25.56 میلیون و

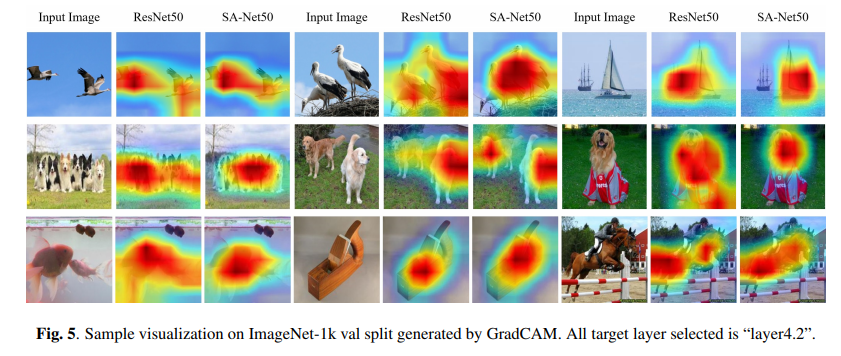
2.76e-3 GFLOPs

در مقابل

4.12 GFLOPs

است و افزایش دقت شبکه بیش از 1.34٪ بوده است. این روش در حال حاضر SOTA می باشد.





این ماژول بهینه ساز بر روی mobile net و همچنین بر روی مسئله چهره استفاده نشده است.

برای معماری پایه

1. Searching for MobileNetV3 - 2019

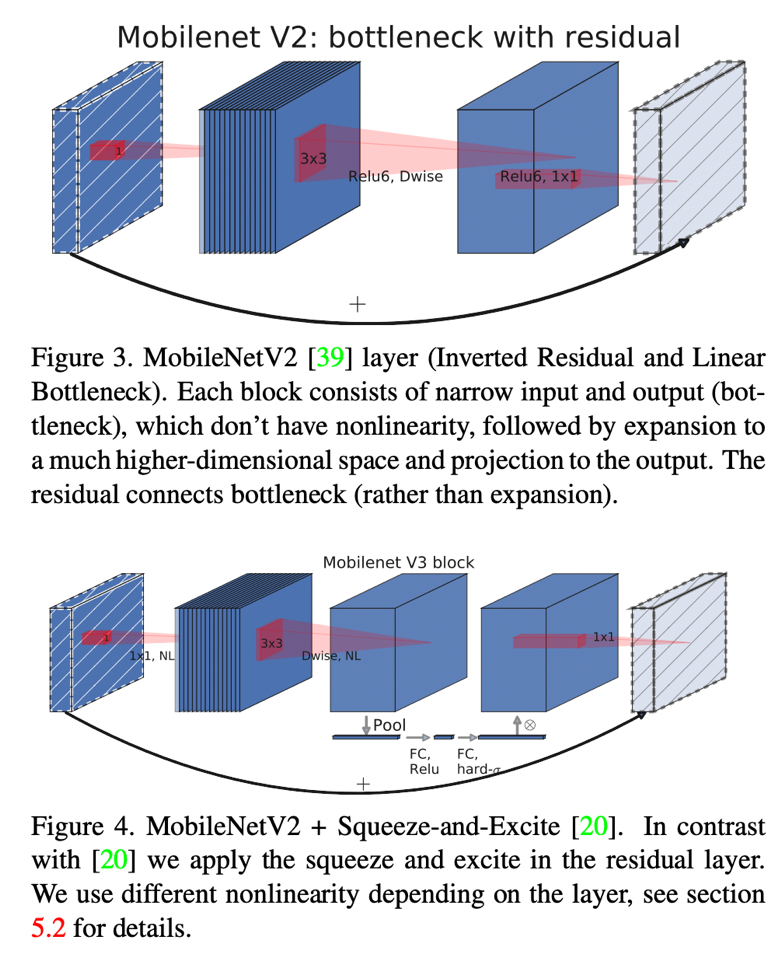
url: <https://arxiv.org/abs/1905.02244>

این مقاله یک طراحی معماری جدید ارائه می دهد به نام. MobileNetV3

دو مدل جدید MobileNet به نام MobileNetV3-Large و MobileNetV3-Small که برای استفاده از منابع کم و زیاد طراحی شده اند.

این مقاله همچنین یک لایه polling جدید پیشنهاد می دهد. Spramial Pyramid Pooling

به طور خلاصه تفاوت بین معماری موبایل نت ورژن ۲ و ورژن ۳ به صورت زیر می باشد:



مزایای این روش:

در دسته بندی به نتایج جدیدی دست یافته است:. MobileNetV3-Large در طبقه بندی ImageNet به میزان 3.2٪ درصد دقیق تر است در حالی که تأخیر را 15٪نسبت به MobileNetV2 کاهش می دهد

. MobileNetV3-Small درحالی که 5٪ تأخیر را نسبت به MobileNetV2 کاهش می دهد ، 4.6٪ درصد دقیق تر است. تشخیص MobileNetV3-Large تقریباً با همان دقت MobileNetV2 در تشخیص COCO به میزان 25٪ سریعتر است

معایب این روش:

تعداد دسته ها (number of classes) ثابت است و پس از آموزش قابل تغییر نیست. ابعاد تصویر چهره ورودی ۲۲۴\*۲۲۴ پیکسل می باشد و برای تصاویر چهره کوچکتر از این مقدار، عملکرد دقیقی ندارد.

# Plan for next week

در حال حاضر مشغول به نگارش فصل ۴ و ۵ پایان نامه هستم و تا هفته آینده ان شا الله نتایج آموزش شبکه را به اشتراک خواهم گذاشت. همچنین در حال نگارش مقاله برای این ایده هستم.