Weekly Report

|  |  |
| --- | --- |
| Project title | Realtime Face Recognition in Unconstraint Environments |
| Report period | Weekly |
| Name | Sajjad Aemmi |
| Date | 2021/05/12 |

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

در گزارش قبلی ۱۱ مقاله جدید در زمینه تشخیص چهره در حالت بی درنگ یا شرایط بدون محدودیت را بررسی کردیم و مزایا و معایب هر روش را جداگانه دسته بندی کردیم.

نتیجه گیری کلی از مقالات فوق به شرح زیر است:

تابع ضرر state of the art در حال حاضر arcFace میباشد که بیشترین دقت را روی resnet برای تسک one shot learning کسب کرده است. این تابع ضرر بر روی mobilenet v3 پیاده سازی نشده است.

روش های super resolution با توجه به استفاده از شبکه Generative در زمان اجرا، اصلا real-time نیستند. همچنین دقت خیلی بالایی بر روی دیتاست های معروف نداشتند.

ایده های جدیدی مانند Transformer ها وجود دارند که به علت چالش زمان پردازشی بالا، در حال حاضر به صورت real time مورد استفاده قرار نمیگیرند.

معماری های mobile net با توجه به معیار نسبت دقت بر سرعت، بهترین گزینه برای راه حل این پایان نامه می باشند. نسخه های مختلفی از این معماری تا کنون منتشر شده است که در ادامه بررسی شده اند.

تمام روش های فوق برای حالتی استفاده شده اند که تعداد دسته ها ثابت باشد. استفاده از شیوه آموزش سایامیس که در گزارش قبل بررسی شد، می تواند به عنوان یک ایده تکمیل کننده به ما کمک کند تا شبکه مورد نظر را برای تعداد دسته های متغیر نیز آموزش دهیم.

# Research activities last week

# 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.

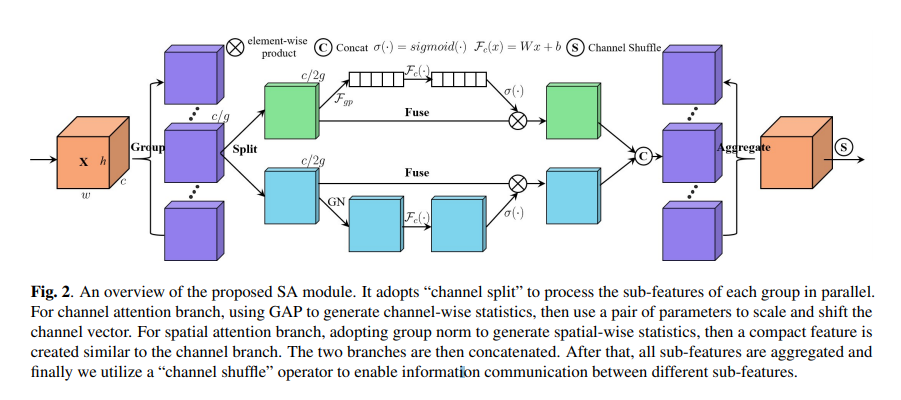
این روش بار محاسباتی بسیار کمی دارد. به عنوان مثال ، پارامترها و محاسبات SA در ResNet50 به ترتیب 300 در مقابل 25.56 میلیون و

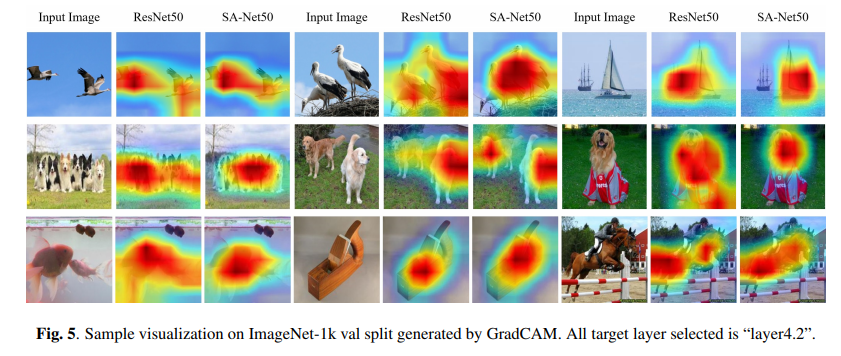
2.76e-3 GFLOPs

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.

# Ultra-thin MobileNet - 2020

url: <https://ieeexplore.ieee.org/document/9031228>

این مقاله برای ایجاد کارآیی و مناسب تر جهت استفاده در embedded systems به صورت real time ، اصلاحاتی را در معماری پایه MobileNet ارائه می دهد. هدف این مقاله این است که به طور قابل توجهی از اندازه ، تعداد پارامترها ، زمان محاسبه در هر فریم کاسته شود. بدون کاهش دقت پایه.

با استفاده از تابع فعال سازی Swish به جای تابع فعال سازی استاندارد ReLU و معرفی روشی به نام پاک کردن تصادفی به جای Drop out ، به دقت خوبی رسیده است. با استفاده از Convolutions قابل جدا شدن به جای Convolutions Depthwise Separable ، تغییر عمق کانال ، انتخاب مقدار ضریب بهینه و حذف برخی از لایه ها با همان شکل خروجی ، بدون افت زیاد در دقت، اندازه مدل را کاهش داده است. معماری بسیار سبک تری را در مقایسه با MobileNet V1 پایه با عنوان MobileNet فوق العاده نازک با اندازه 3.9 مگابایت طراحی کرده اند که در پردازنده های تلفن همراه به صورت بی درنگ با حافظه و قدرت محدود قابل استفاده است.

# MobiExpressNet: A Deep Learning Network for Face Expression Recognition on Smart Phones – 2020

url: <https://ieeexplore.ieee.org/document/9042973>

این مقاله کنفرانسی می باشد. این مقاله یک مدل جدید یادگیری عمیق سبک با نام MobiExpressNet ، برای تشخیص چهره معرفی می نماید.

مدل mobile net معمولی از depthwise separable convolutions استفاده می نماید که نسبت به کانولوشن عادی، بسیار زمان پردازشی کمتری دارد. این مدل پیچیدگی depthwise separable convolutions را محدودتر کرده است ، و با downsampling سعی کرده لایه های شبکه را کمتر کرده تا اندازه مدل را بسیار کوچک نگه دارد.

اندازه مدل MobiExpressNet بیش از 5 برابر کوچکترین مدل MobileNet است که باعث می شود این مدل کاملا real time عمل نماید. اما در مقابل دقت آن کاهش پیدا کرده است.

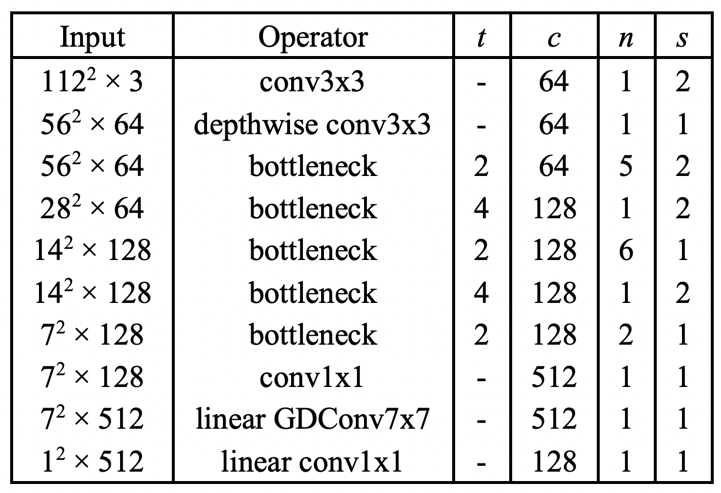
.

# MobileFaceNets: Efficient CNNs for Accurate Real- Time Face Verification on Mobile Devices - 2019

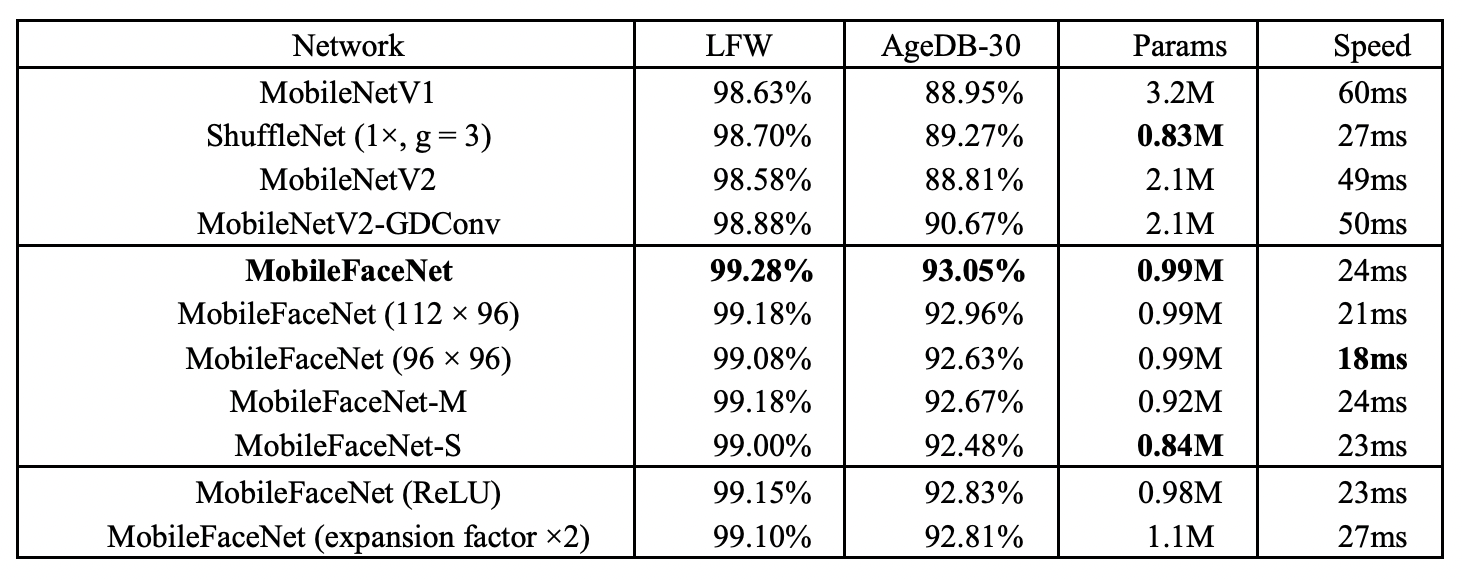
url: <https://arxiv.org/pdf/1804.07573.pdf>

این مقاله مدل MobileFaceNets را ارائه می دهد که کمتر از 1 میلیون پارامتر دارد و بطور خاص برای تشخیص چهره به صورت realtime با دقت بالا در دستگاه های تلفن همراه و embedded device ها طراحی شده است. نقاط ضعف شبکه mobilenet توسط MobileFaceNets برطرف شده است. این مدل بیش از 2 برابر سرعت بیش از MobileNetV2 دارد. پس از آموزش ArcFace بر روی MS-Celeb-1M بدون نویز، این مدل با اندازه 4.0 مگابایت به دقت 99.55٪ در LFW و 92.59٪ در MegaFace رسیده است.

معماری شبکه:



مقایسه نتایج با روش های مشهور دیگر:

****

نقاط قوت:

این روش با پیشرفته ترین مدلهای بزرگ CNN با اندازه صدها مگابایت قابل مقایسه است. زمان پردازش بر روی تلفن همراه 18 میلی ثانیه است.

نقاط ضعف:

ابعاد تصویر چهره ورودی ۱۱۲\*۱۱۲ پیکسل می باشد و برای تصاویر چهره کوچکتر از این مقدار، عملکرد دقیقی ندارد.

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

1. Searching for MobileNetV3 - 2019

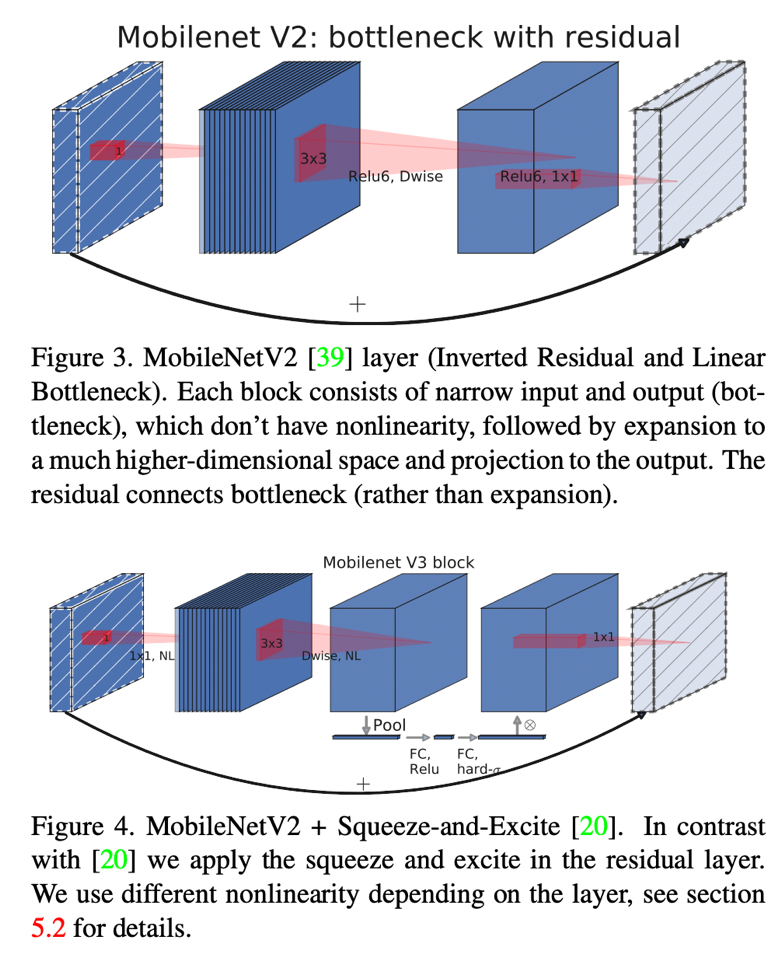
url: <https://arxiv.org/abs/1905.02244>

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.

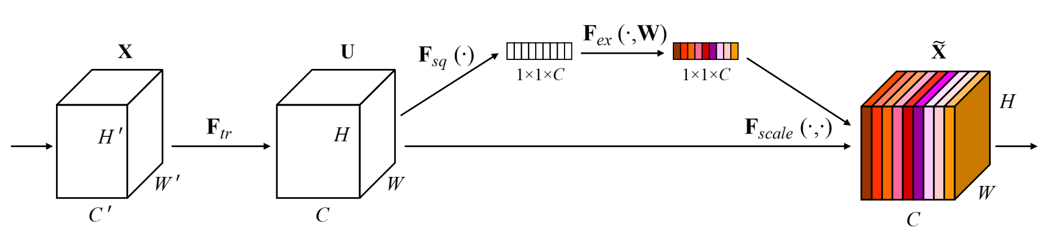
1. Squeeze-and-Excitation Networks - 2019

url: <https://arxiv.org/pdf/1709.01507.pdf>

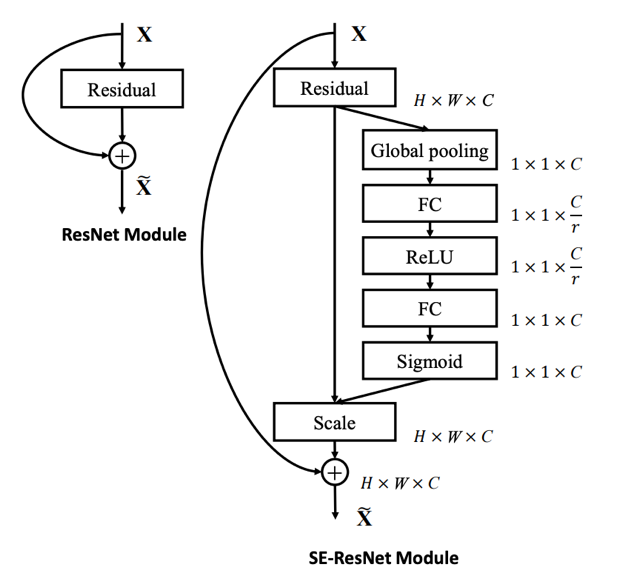
در این معماری بر خلاف cnn معمولی که از لایه های کانولوشنی استفاده می نماید، بلوک هایی با نام

“Squeeze-and-Excitation” (SE) block

وجود دارند که از کانولوشن های با ابعداد ۱\*۱ استفاده می کند که تعداد پارامترها و بار محاسباتی کمی دارند. اما با افزودن این بلوک ها به معماری شبکه هایی مانند resnet و mobilenet دقت این شبکه ها افزایش پیدا کرده است.



برای مثال در شکل زیر مقایسه بین بوک resnet عادی و SE-resnet را مشاهده می نمایید.



مزایای این روش:

افزایش دقت معماری های مشهور به کمک افزودن بلوک های SE

معایب این روش:

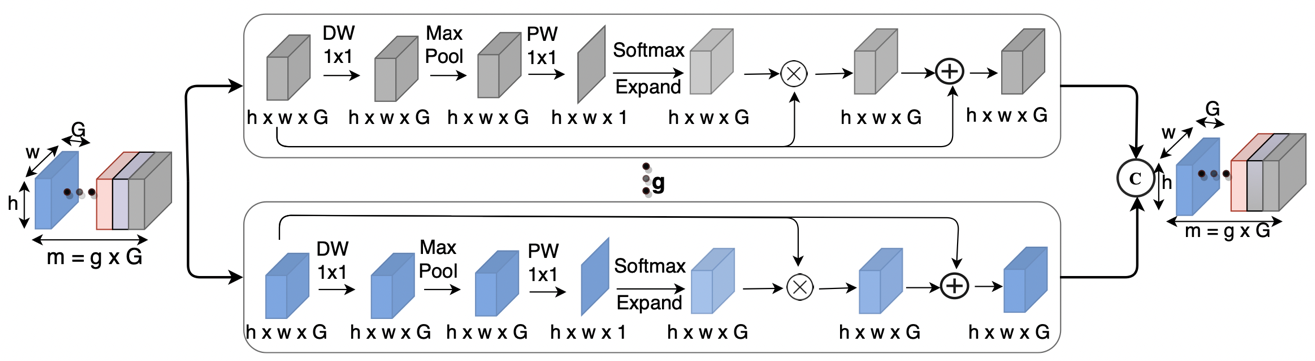
افزایش بار محاسباتی در مقایسه با شبکه های قبلی. برای مثال تعداد پارامتر های موبایل نت از ۴.۲ میلیون به ۴.۷ میلیون رسیده است.

عدم آزمایش بر روی دیتاست های مربوط به مسائل تشخیص چهره

1. ULSAM: Ultra-Lightweight Subspace Attention Module for Compact Convolutional Neural Networks – 2020

url: <https://arxiv.org/pdf/2006.15102.pdf>

استفاده از لایه های attention می تواند کارایی یک مدل شبکه عصبی عمیق را افزایش دهد. اما از طرفی بار محاسباتی سنگینی را متحمل می شود. دراین مقاله از یک مکانیزم لایه های attention سبک با هزینه محاسباتی کم صحبت شده است. و سپس از آن در معماری MobileNet-V1 و MobileNet-V2 استفاده شده است. ساختار کلی بلوک توجه در شکل زیر قابل مشاهده است.



مزایای این روش:

۱۳ تا ۲۵ درصد کاهش هزینه محاسباتی داشته است. و ۱ درصد افزایش دقت در MobileNet-V2 داشته است.

معایب این روش:

بر روی دیتاست چهره کار نشده است. و معماری MobileNet-V2 نسبت به MobileNet-V3 قدیمی تر می باشد.

1. Lightweight FaceNet Based on MobileNet – 2021

url: <https://www.scirp.org/journal/paperinformation.aspx?paperid=104599>

این مقاله معماری faceNet که یکی از بهترین و قوی ترین مقاله ها در زمینه تشخیص چهره بوده، بر أساس معماری MobileNet تغییر داده و سبک تر و سریع تر شده.

FaceNet مدلی است که در سالهای اخیر در تشخیص چهره استفاده می شود. FaceNet از معماری GoogLeNet استفاده می کند که دقت بالایی در تشخیص چهره دارد. با این حال ، ساختار شبکه آن بیش از حد بزرگ است که باعث می شود FaceNet با سرعت کم کار کند.

بنابراین برای بهبود سرعت، بدون تأثیر بر دقت تشخیص FaceNet ، این مقاله مدل سبک FaceNet را بر اساس MobileNet پیشنهاد می کند. این مدل مشابه موبایل نت با استفاده از deep separable convolutions ، محاسبه کلی شبکه را کاهش می دهد.

مزایای این روش:

مدل در مجموعه داده های CASIA-WebFace و VGGFace2 آموزش داده شده و بر روی مجموعه داده LFW آزمایش شده است. نتایج نشان می دهد که این مدل ضمن دقت بالا، پارامترهای شبکه را تا حد زیادی کاهش می دهد. این مدل همچنین می تواند تشخیص چهره را روی شخص خاصی در ویدیو انجام دهد.

معایب این روش:

این مدل بر أساس معماری موبایل نت ورژن ۱ طراحی شده. اما موبایل نت ورژن ۲ و ۳ جدید تر هستند.

دقت نسبت به مدل اصلی facenet با اندکی کاهش همراه بوده است. تعداد دسته ها (number of classes) ثابت است و پس از آموزش قابل تغییر نیست.

# Problems encountered

# جمع بندی:

# 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.

# دو ایده جدید در این زمینه می تواند عملی شود که تا کنون مورد استفاده قرار نگرفته است:

# 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.

# استفاده از SE block که در شبکه Squeeze-Net استفاده شده ست. و ادغام آن با mobile net به منظور بهره گیری از مزایای هر دو شبکه به صورت همزمان.

# میدانیم که mobile net دقت بالاتری دارد. اما Squeeze-Net از لایه توجه استفاده کرده و نسبت به mobile net سبک تر می باشد. مشابه ایده اول، در نهایت این مدل را به روش سایامیس و تابع ضرر arc face آموزش می دهیم.

# 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.

# Plan for next week

با مطالعه مقاله های نزدیک و مرتبط با موضوع پایان نامه، مسیر پیش رو، روشن تر شد. در هفته آینده بر روی تکمیل فصل روش پیشنهادی کار تمرکز کرده و ایده مورد نظر را به همراه جزییات نگارش خواهم کرد.