Weekly Report

|  |  |
| --- | --- |
| Project title | Realtime Face Recognition in Unconstraint Environments |
| Report period | Weekly |
| Name | Sajjad Aemmi |
| Date | 2021/04/29 |

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

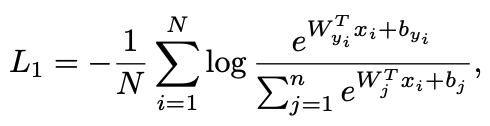
# Research activities last week

1. CosFace: Large Margin Cosine Loss for Deep Face Recognition – 2018

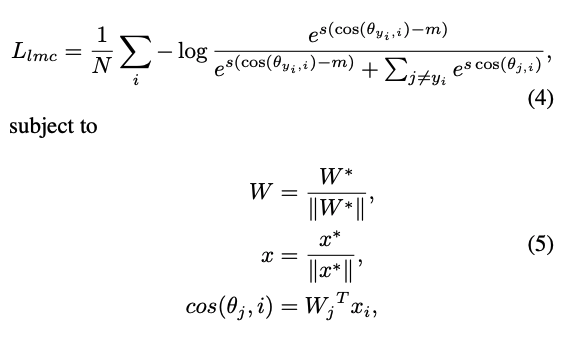
url: <https://arxiv.org/pdf/1801.09414.pdf>

این مقاله یک تابع ضرر جدید تعریف کرده است که بر اساس کمینه کردن کسینوس زاویه بین بردارهای ویژگی در فضای n بعدی عمل می نماید. این تابع ضرر که در مقایسه با softmax دقت دسته بندی بیشتری دارد، به گونه ای عمل می نماید که فاصله داده های درون دسته ای را کمتر و فاصله داده های برون دسته ای را بیشتر نماید.

تابع ضرر softmax:



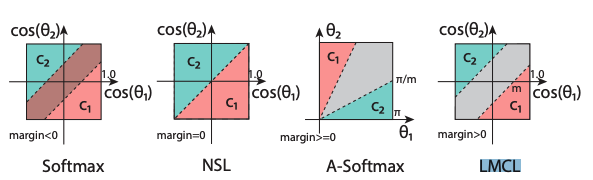
تابع ضرر CosFace:



مزایای این روش:

دقت ۹۹.۷۳ روی دیتاست LFW و دقت ۹۷.۶ بر روی دیتاست YTF

ایجاد حاشیه امن بین مرز دسته ها مشابه SVM



معایب این روش:

این مقاله تابع ضرر پیشنهادی خود را بر روی یک CNN با ۶۴ لایه (شبیه ResNet) آموزش داده است. که real-time نمی باشد.

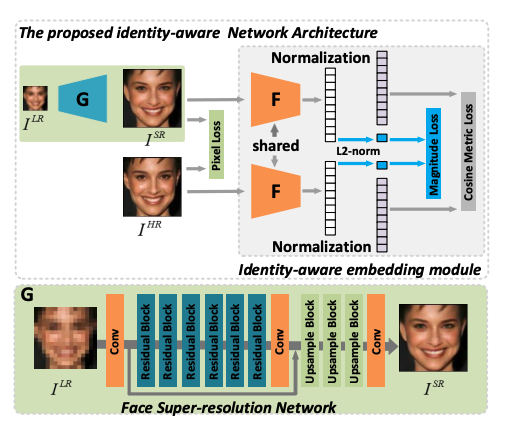
دقت این روش بر روی دیتاست MegaFace Challenge 1 - MF1 برابر ۷۷ درصد است.

1. Identity-Aware Face Super-Resolution for Low-Resolution Face Recognition - 2020

url: <https://ieeexplore.ieee.org/document/9072532>

این مقاله یک روش با تمرکز بر روی داده های Low-Resolution و تبدیل آن ها به تصاویر چهره با کیفیت ارائه داده است.

با استفاده از شبکه ی Generative یک تصویر چهره با ابعاد ۳۲\*۳۲ پیکسل را به یک تصویر چهره با کیفیت بالا تبدیل می کند و سپس توسط یک شبکه CNN که ویژگی های مبتنی بر هویت استخراج می نماید، شبکه Generative را آموزش می دهد. برای دستیابی به تصاویر Super-Resolution از معماری زیر استفاده می کند.



مزایای این روش:

نسبت به تغییر اندازه تصویر به روش Bicubic افزایش دقت ۱۰ درصدی در دیتاست LFW داشته است. و در مواردی که اندازه تصویر چهره ۳۲\*۳۲ یا ۱۶\*۱۶ است، عملکردی قابل قبولی دارد.

معایب این روش:

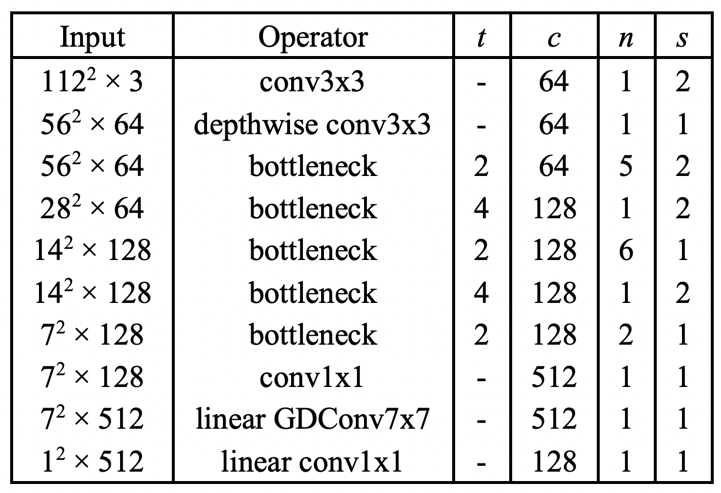
بر روی دیتاست CelebA دقت ۴۱ درصد گرفته است. و با توجه به استفاده از شبکه Generative در زمان اجرا، این روش اصلا real-time نیست.

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

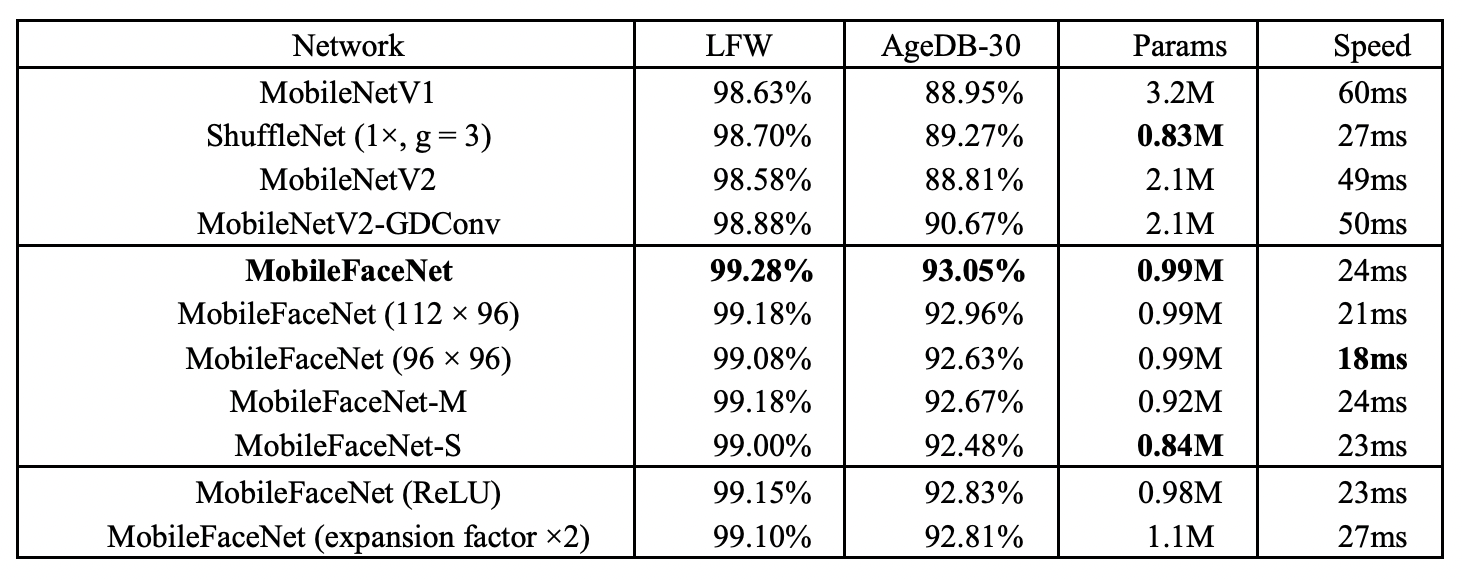
url: <https://arxiv.org/pdf/1804.07573.pdf>

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.

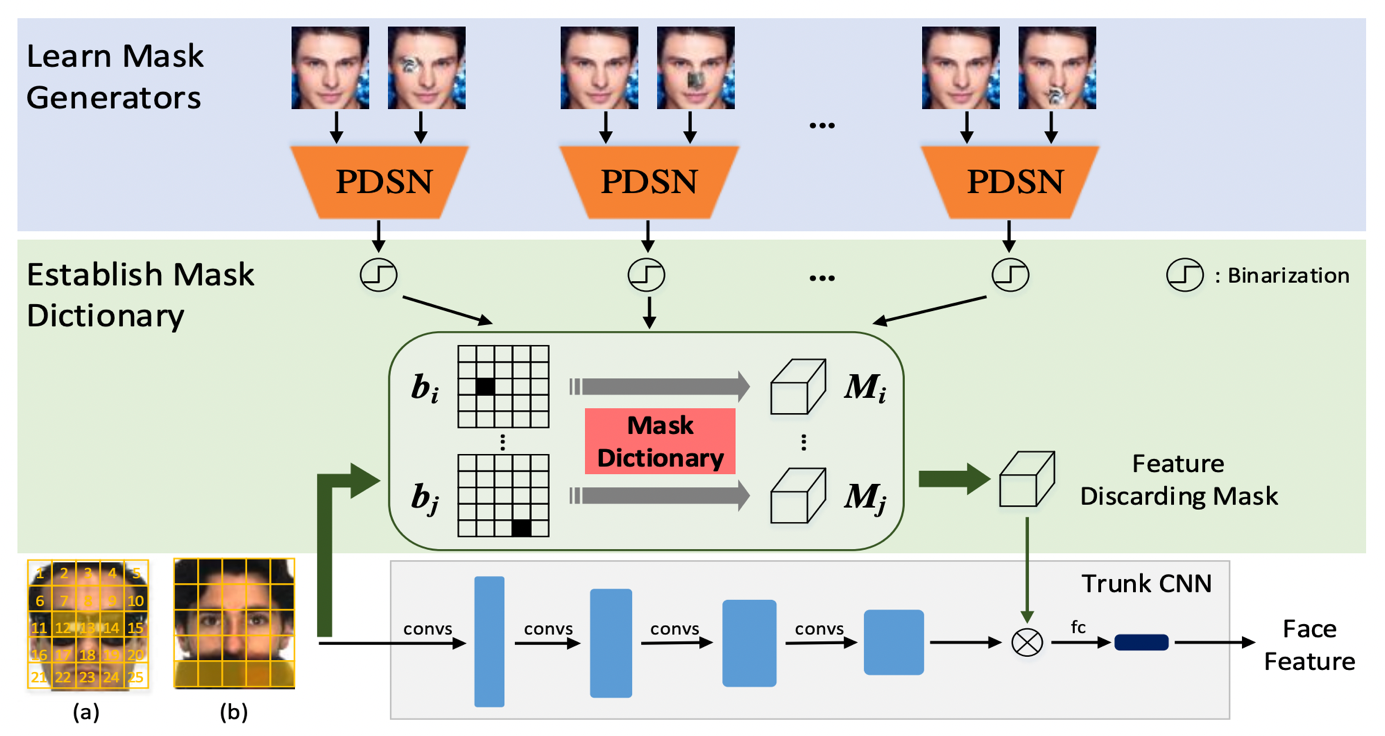
# Occlusion Robust Face Recognition Based on Mask Learning With Pairwise Differential Siamese Network – 2019

url:<https://www.researchgate.net/publication/330308706_Small_sample_face_recognition_algorithm_based_on_novel_Siamese_network>

یکی از راه حل های حل مشکل تعداد متغیر دسته ها، استفاده از روش آموزش شبکه سیامیس است. در این روش لایه softmax بالای شبکه را حذف میکنیم. و آخرین لایه ماقبل آن که ویژگی های استخراج شده در آن قرار دارد را مستقیما مورد استفاده قرار می دهیم. داده های آموزش را به صورت دو به دو و pair جدا میکنیم. به صورتی که بعضی از pair ها متعلق به یک شخص یکسان (لیبل یک) و بعضی ها متعلق به دو شخص مختلف می باشد (لیبل صفر).

بدین ترتیب شبکه طوری آموزش می یابد که ویژگی هایی استخراج نماید که برای هویت یکسان، نزدیک به هم، و برای هویت های متفاوت، دور از هم باشد.

همچنین با ایجاد داده های augment شده در شرایط انسداد مختلف، شبکه را rebust تر کرده است.



مزایای این روش:

قابلیت تغییر تعداد دسته ها پس از آموزش شبکه. قابل استفاده در شرایطی که داده های آموزش کافی در اختیار نداریم. قدرت بالا در شرایط انسداد

معایب این روش:

دقت پایین تر در مقایسه با cosFace و ArcFace – بر روی LFW دقت ۹۹.۲۰ درصد

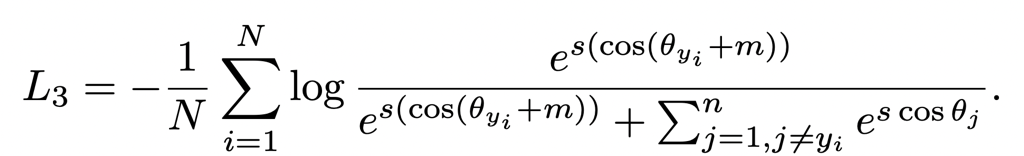
توابع ضرر cosFace و ArcFaceبسیار بهتر عمل کردند

عدم سرعت بالای الگوریتم و توجه به realtime

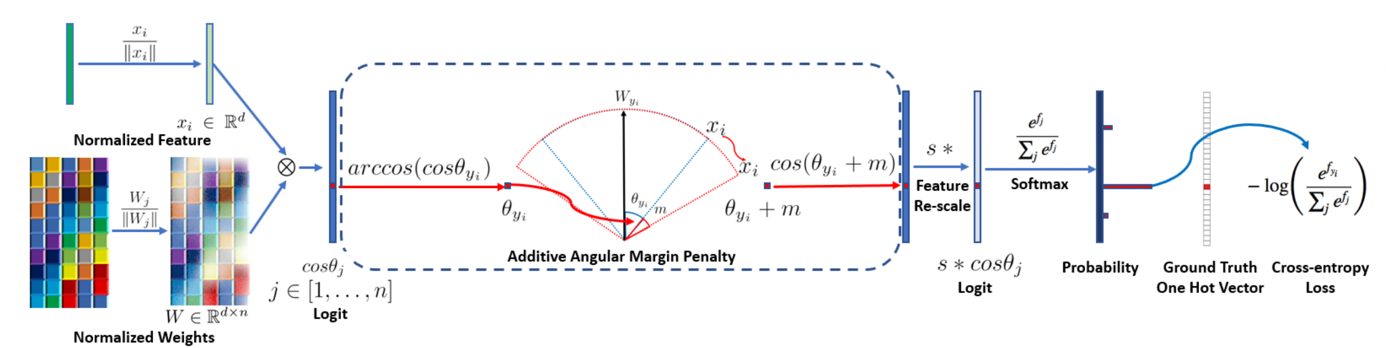
1. ArcFace: Additive Angular Margin Loss for Deep Face Recognition – 2019

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:

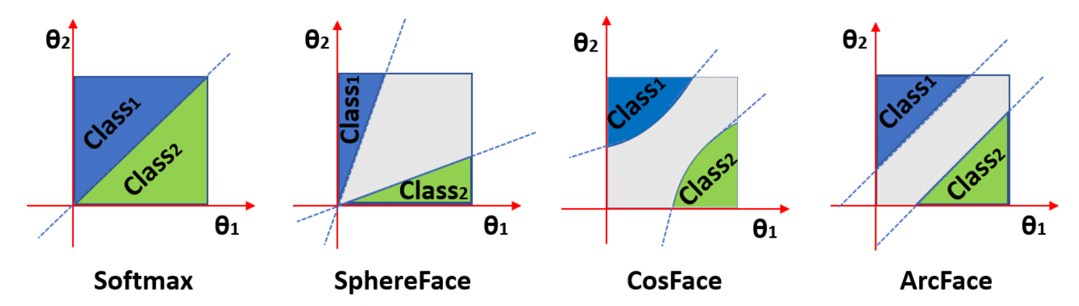


خلاصه این روش در شکل زیر آمده است:



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

1. Searching for MobileNetV3 - 2019

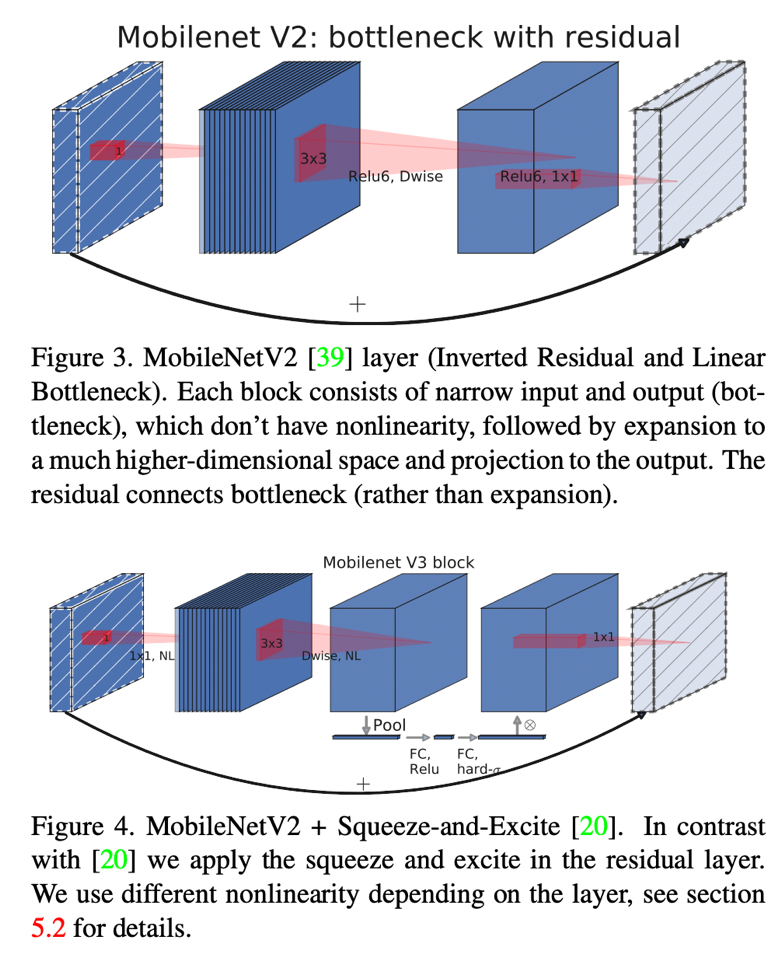
url: <https://arxiv.org/abs/1905.02244>

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

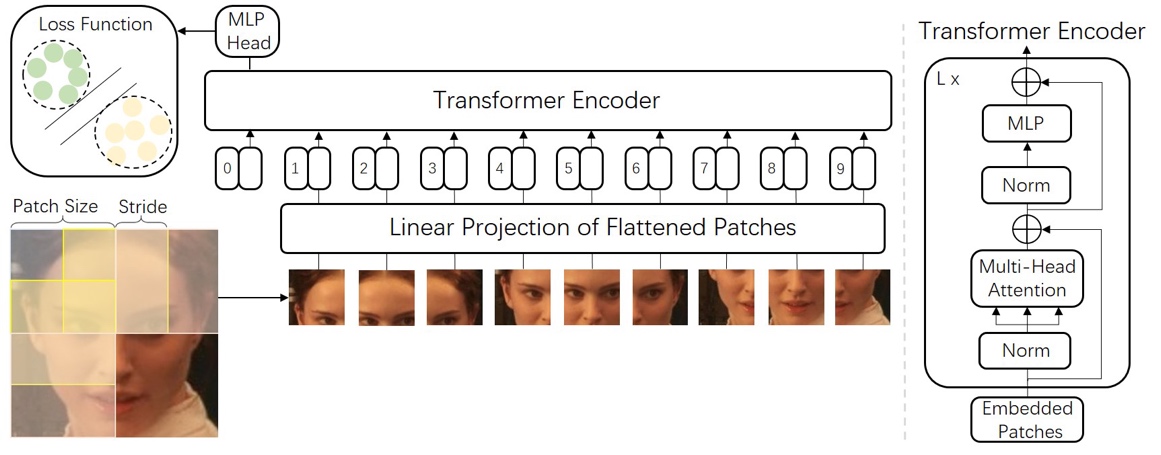
معایب این روش:

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. Face Transformer for Recognition – 2021

url: <https://arxiv.org/pdf/2103.14803v2.pdf>

این مقاله از ایده استفاده از Transformer بهره برده است که تا کنون بیشتر در مسائل ترتیبی مانند NLP مورد استفاده قرار میگرفت. به این شکل که تصویر به patch های مختلف تقسیم می شود و سپس با حفظ ترتیب پچ ها، وارد شبکه Transformer می شوند. در خروجی این شبکه یک mlp قرار دارد که عملیات دسته بندی را انجام می دهد.



مزایای این روش:

بسیار جدید و تازه است

بر خلاف cnn که نگاه local به تصویر داشت. این شبکه نگاه global نیز به تصویر دارد.

معایب این روش:

هنوز دقت آن از شبکه های cnn بیشتر نشده است. و یک شبکه مانند resNet100 دارای دقت بالاتری از این روش می باشد.

این روش به علت استفاده از Transformer سرعت بسیار پایینی دارد و مناسب اجرا روی پردازنده تلفن همراه نیست.

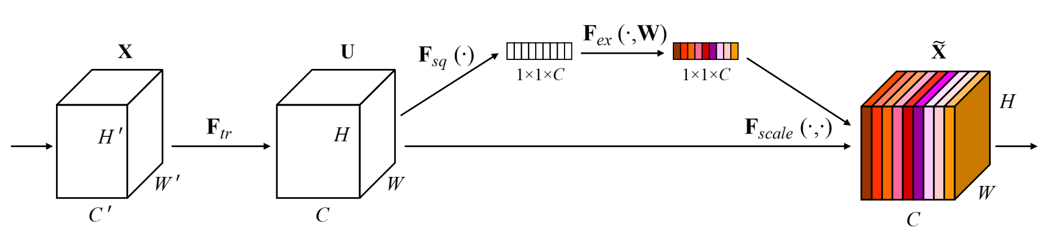
1. Squeeze-and-Excitation Networks - 2019

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

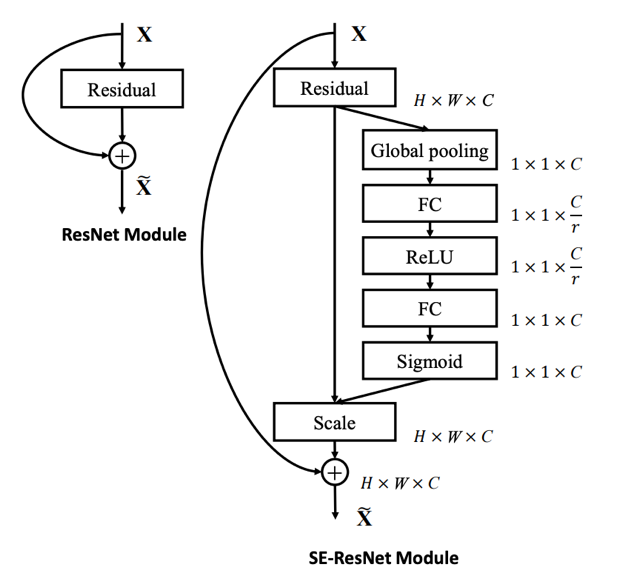
در این معماری بر خلاف cnn معمولی که از لایه های کانولوشنی استفاده می نماید، بلوک هایی با نام

“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:

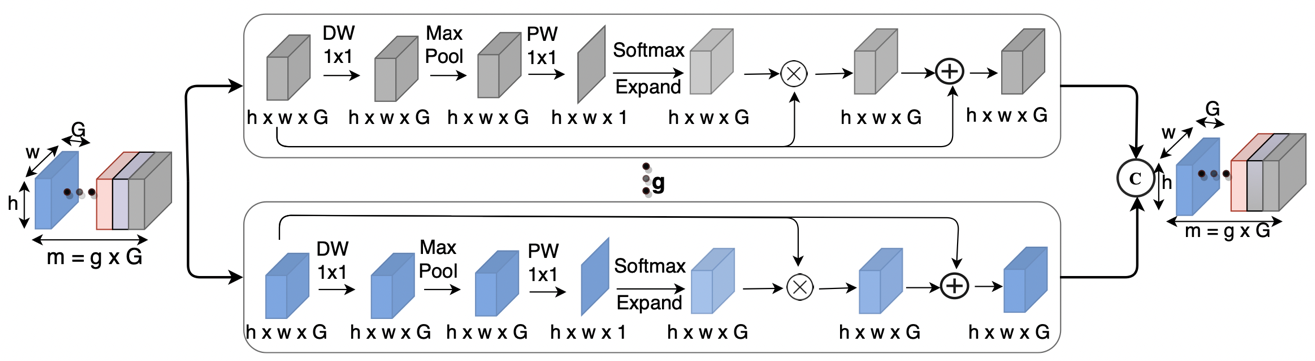
افزایش بار محاسباتی در مقایسه با شبکه های قبلی. برای مثال تعداد پارامتر های موبایل نت از ۴.۲ میلیون به ۴.۷ میلیون رسیده است.

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. Efficient Low-Resolution Face Recognition via Bridge Distillation – 2020

url: <https://ieeexplore.ieee.org/document/9098036>

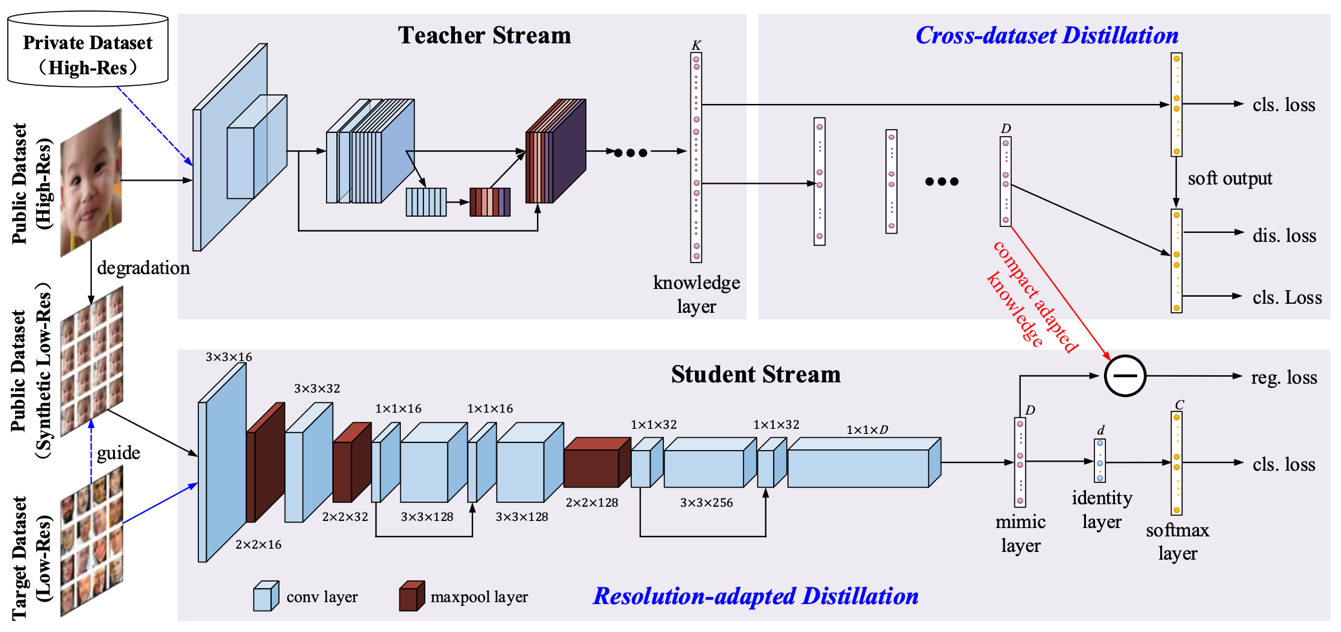
ایده این مقاله طراحی یک معماری مبتنی بر مدل teacher و student می باشد.

ابتدا مدل teacher بر روی یک دیتاست اولیه high-resolution آموزش داده می شود. سپس این مدل pre-trained بر روی یک دیتاست high-resolutionدیگر آموزش داده می شود تا بتواند ویژگی های discriminative خوبی استخراج نماید.

در نهایت یک مدل student آموزش داده می شود که از بردار ویژگی تولید شده توسط مدل teacher برای تشخیص بهتر

در مرحله آموزش هر دو شبکه استفاده می شوند.

در مرحله آزمایش فقط از شبکه student استفاده می شود.



مزایای این روش:

نتایج تجربی نشان می دهد که مدل student در شناسایی چهره های با وضوح پایین با فقط 0.21 میلیون پارامتر و حافظه 0.057MB عملکرد مناسبی دارد. سرعت آن به ترتیب در GPU ، CPU و تلفن همراه به 14705 ، 934 و 763 چهره در ثانیه می رسد.

معایب این روش:

نیاز به چندین مجموعه داده آموزشی بسیار زیاد

دقت حدود ۸۰ درصد بر روی مجموعه داده آزمایشی LFW

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

# 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.