به نام خدا

دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

درس شبکه‌های عصبی و یادگیری عمیق

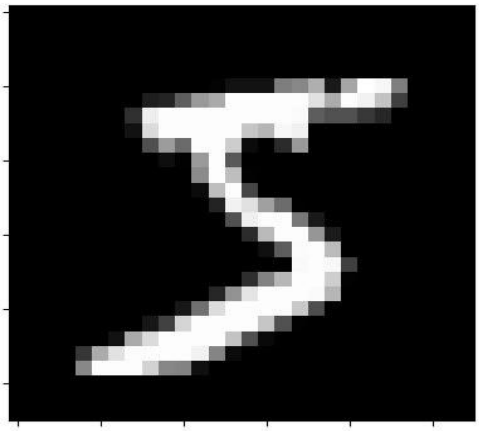
تمرین امتیازی

امیرعباس رضا سلطانی	**610399205**
نیما نیرومند	610399199

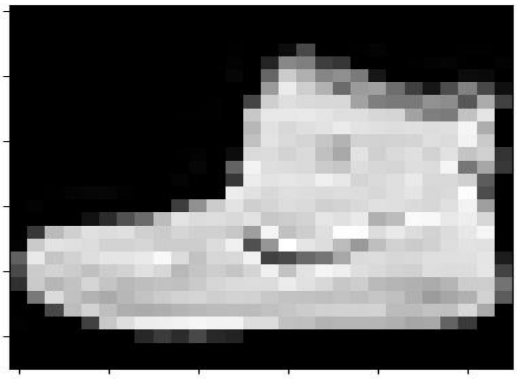
پرسش 1. تولید برچسب به کمک خوشه بندی

۱-۱. دادگان

به کمک کتابخانه keras دو مجموعه داده را لود می کنیم. ابعاد تصویر در هر دیتاست، 28 در 28 است. برای هر دیتاست، تعداد 60000 داده آموزش و 10000 داده تست داریم.



تصویر1. نمونه داده دیتاست mnist

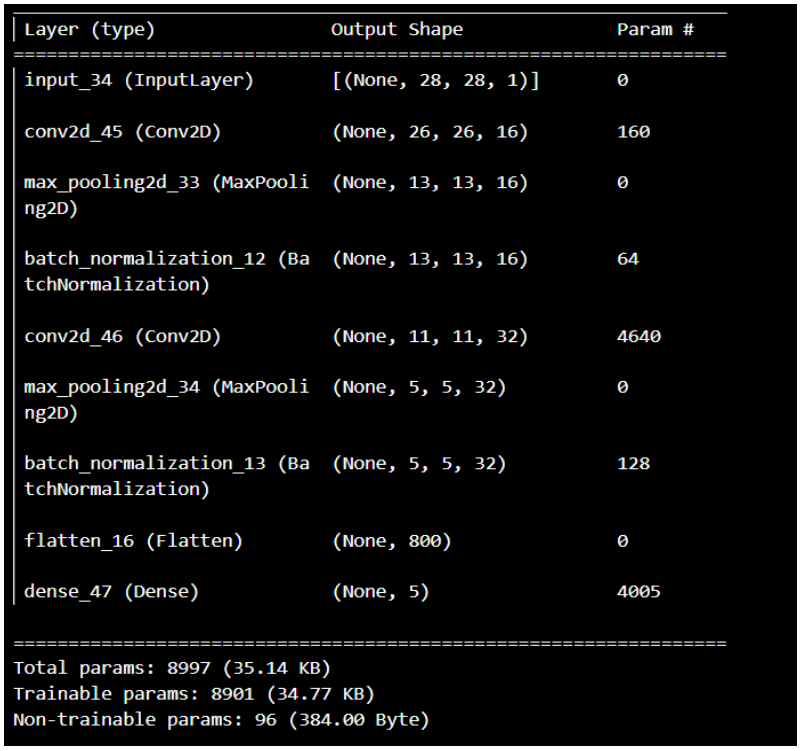


تصویر2. نمونه داده دیتاست fmnist

25 درصد از داده های هردو دیتاست را به عنوان validation در نظر گرفته و مقادیر هر تصویر را نرمالایز می کنیم (با تقسیم بر 255) و به بازه بین صفر تا یک انتقال می دهیم.

2-1. شبکه مورد استفاده

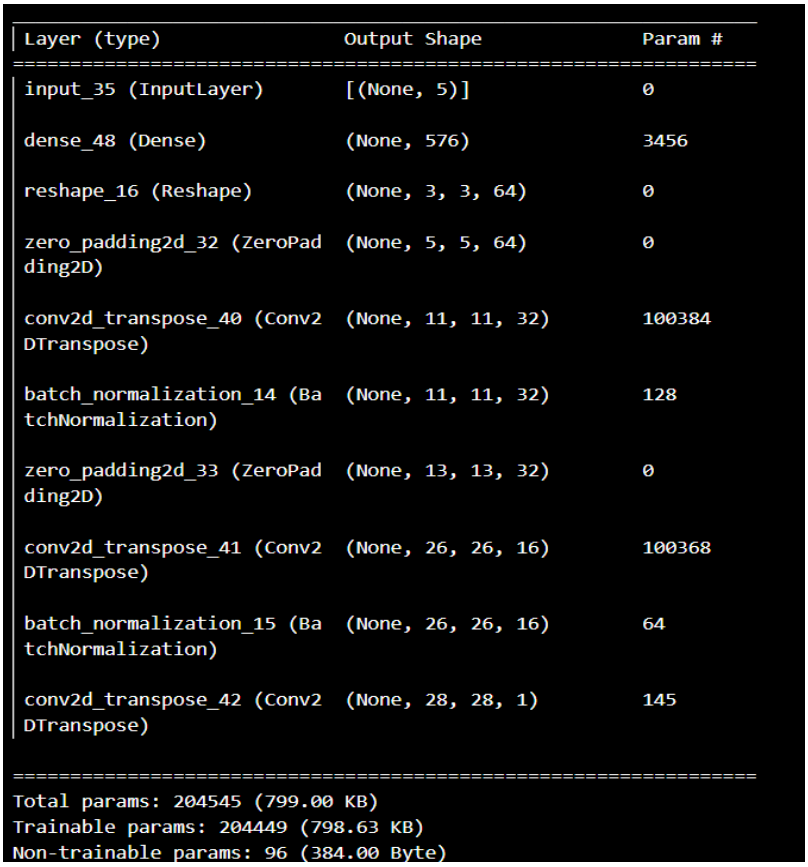
از کتابخانه keras برای پیاده سازی شبکه استفاده می کنیم.



Layer (type)	Output Shape	Param #
input_34 (InputLayer)	[(None, 28, 28, 1)]	0
conv2d_45 (Conv2D)	(None, 26, 26, 16)	160
max_pooling2d_33 (MaxPooling2D)	(None, 13, 13, 16)	0
batch_normalization_12 (BatchNormalization)	(None, 13, 13, 16)	64
conv2d_46 (Conv2D)	(None, 11, 11, 32)	4640
max_pooling2d_34 (MaxPooling2D)	(None, 5, 5, 32)	0
batch_normalization_13 (BatchNormalization)	(None, 5, 5, 32)	128
flatten_16 (Flatten)	(None, 800)	0
dense_47 (Dense)	(None, 5)	4005

Total params: 8997 (35.14 KB)
Trainable params: 8901 (34.77 KB)
Non-trainable params: 96 (384.00 Byte)

تصویر 3. شبکه encoder



Layer (type)	Output Shape	Param #
input_35 (InputLayer)	[(None, 5)]	0
dense_48 (Dense)	(None, 576)	3456
reshape_16 (Reshape)	(None, 3, 3, 64)	0
zero_padding2d_32 (ZeroPadding2D)	(None, 5, 5, 64)	0
conv2d_transpose_40 (Conv2DTranspose)	(None, 11, 11, 32)	100384
batch_normalization_14 (BatchNormalization)	(None, 11, 11, 32)	128
zero_padding2d_33 (ZeroPadding2D)	(None, 13, 13, 32)	0
conv2d_transpose_41 (Conv2DTranspose)	(None, 26, 26, 16)	100368
batch_normalization_15 (BatchNormalization)	(None, 26, 26, 16)	64
conv2d_transpose_42 (Conv2DTranspose)	(None, 28, 28, 1)	145

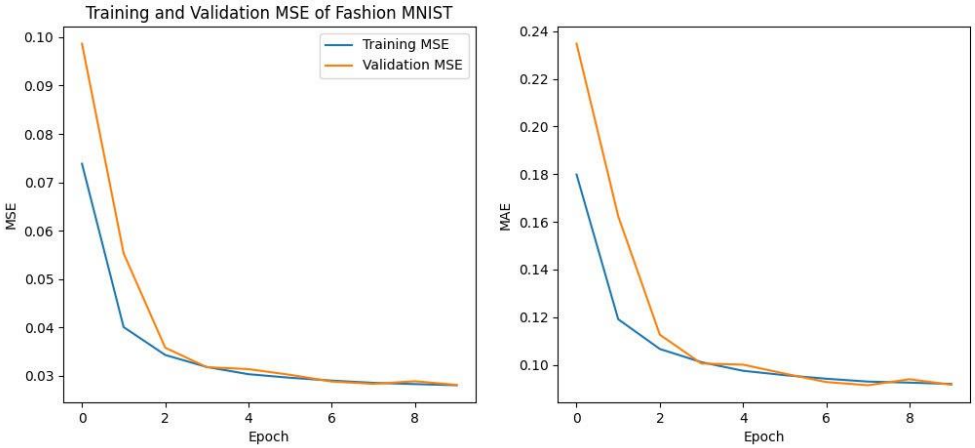
Total params: 204545 (799.00 KB)
Trainable params: 204449 (798.63 KB)
Non-trainable params: 96 (384.00 Byte)

تصویر 4. شبکه decoder

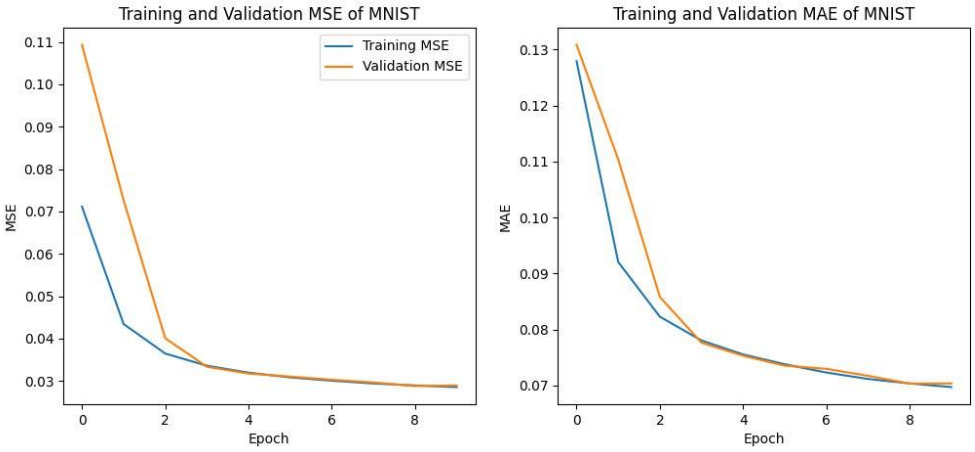
در این autoencoder از فضای latent با طول 5 استفاده شده است.

3-1. آموزش شبکه

مدل با ساختار بیان شده در بخش قبل را یکبار برای داده های mnist و یکبار برای داده های fmnist آموزش می دهیم. برای آموزش از بهینه ساز adam با 0.001 learning rate استفاده می کنیم. مدل را با 256 batch size و 10 epochs آموزش می دهیم.



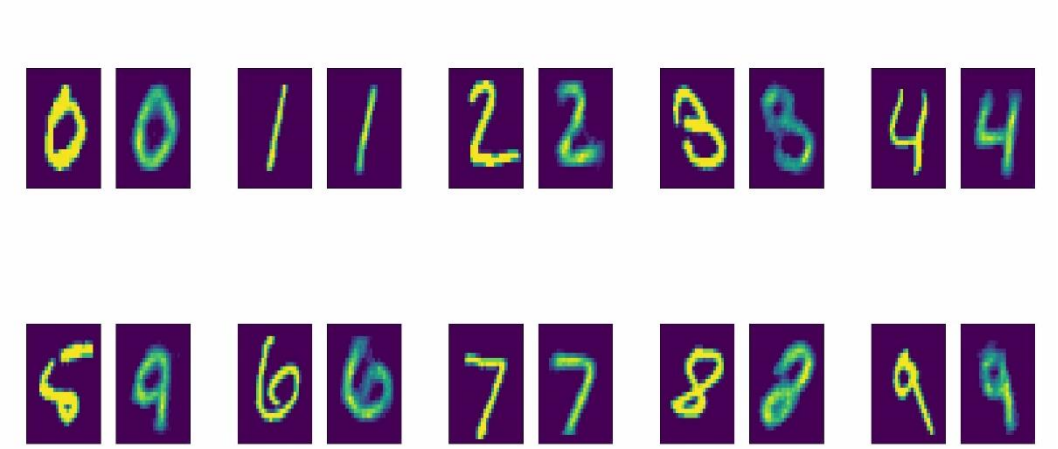
تصویر5. روند تغییر MSE و MAE برای داده های fmnist



تصویر6. روند تغییر MSE و MAE برای داده های mnist

جدول1. نتایج مدل روی داده های تست

	MSE	MAE
Mnist	0.029	0.070
FMnist	0.028	0.094



تصویر7. مقایسه تصویر اصلی و تولید شده توسط شبکه از دیتاست mnist (تصویر سمت چپ عدد واقعی و تصویر سمت راست عدد خروجی شبکه است)

تصویر8. مقایسه تصاویر اصلی و تولید شده توسط شبکه از دیتاست fashion mnist (تصویر سمت چپ fashion واقعی و تصویر سمت راست fashion خروجی شبکه است)

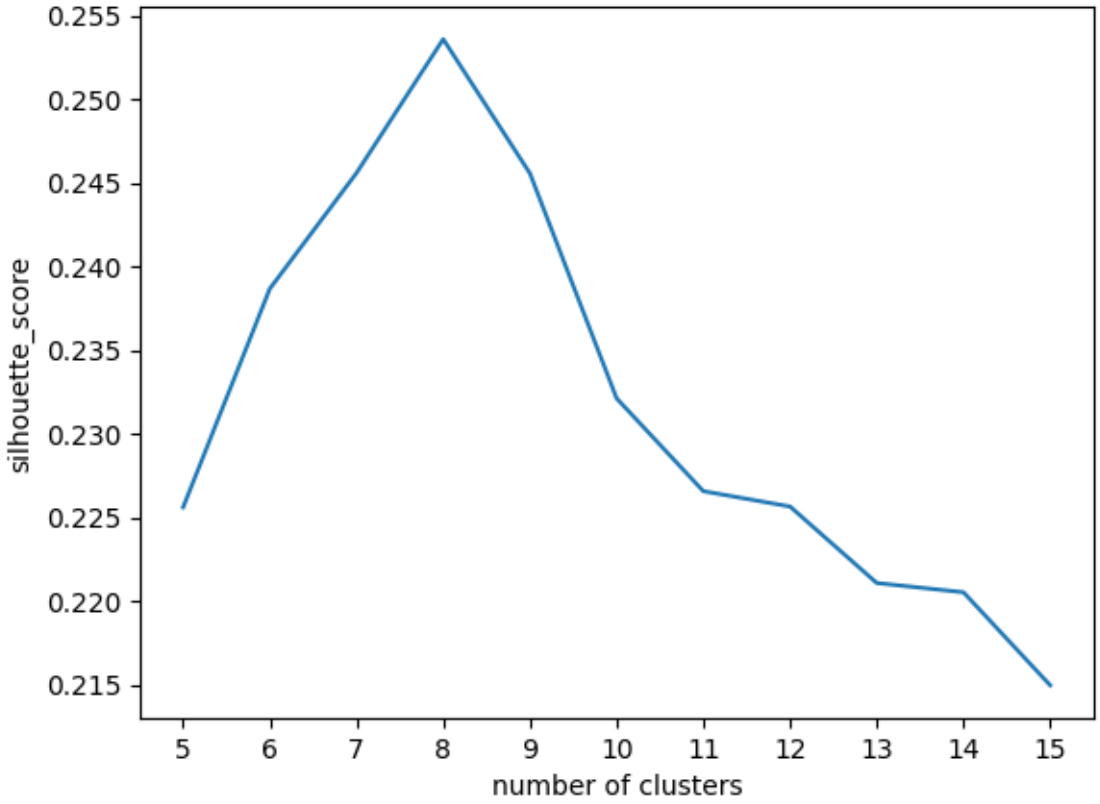
4-1. خوشه بندی

بخش encoder شبکه را پس از آموزش روی هر دیتاست جدا می کنیم. خروجی آن را به برداری با طول 5 تبدیل می کنیم. الگوریتم kmeans را برای تعداد خوشه های 5 تا 15 اجرا می کنیم و بهترین تعداد خوشه را بر اساس معیار silhouette score انتخاب می کنیم.

silhouette score برای یک داده معین از رابطه زیر بدست میاید.

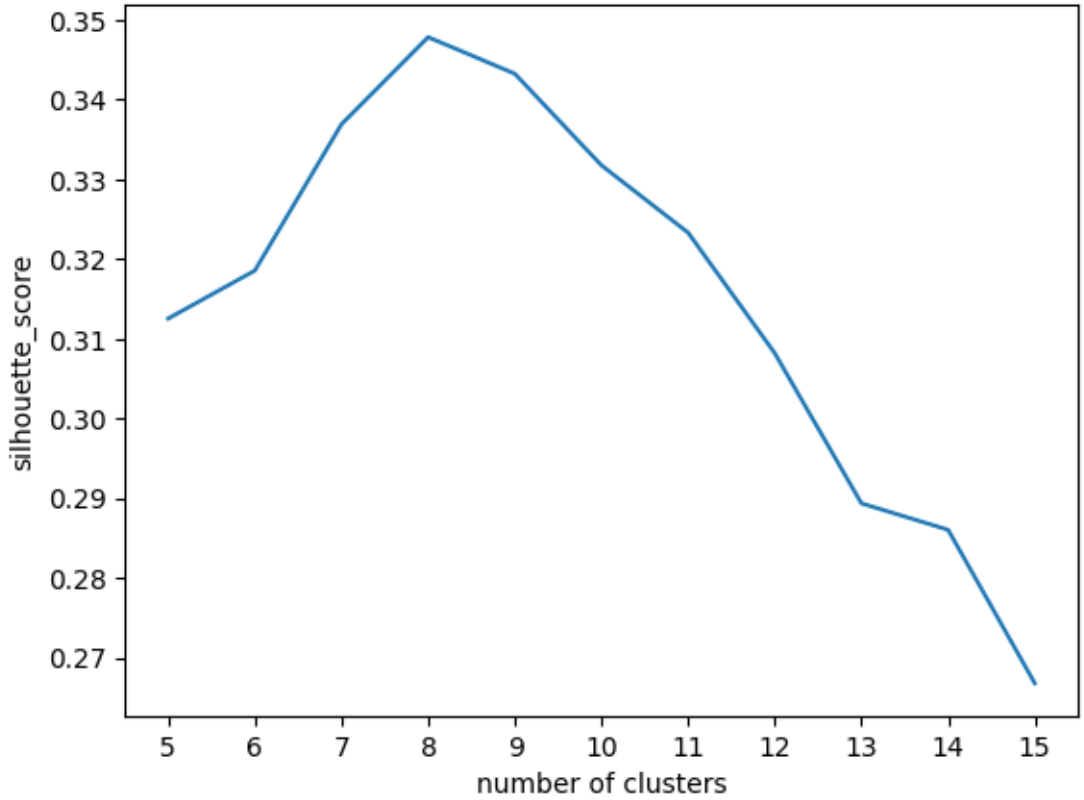
$$\frac{b-a}{max(a,b)}$$

حاصل عددی بین -1 و 1 است. b میانگین فاصله داده از نزدیکترین cluster همسایه و a میانگین فاصله داده از داده های درون همان cluster است. بنابراین میانگین این معیار برای تمام داده ها وقتی ماکزیموم می شود که داده ها درون یک کلاستر به یکدیگر در نزدیکترین حالت ممکن و از داده های کلاسترهای دیگر در دورترین حالت ممکن باشند. در این بخش overall Silhouette Score که میانگین Silhouette Score برای همه داده ها است را محاسبه کرده ایم. بهترین تعداد خوشه، دارای بیشترین overall Silhouette Score است.



تصویر9. نمودار silhouette score براساس تعداد cluster برای دیتاست mnist

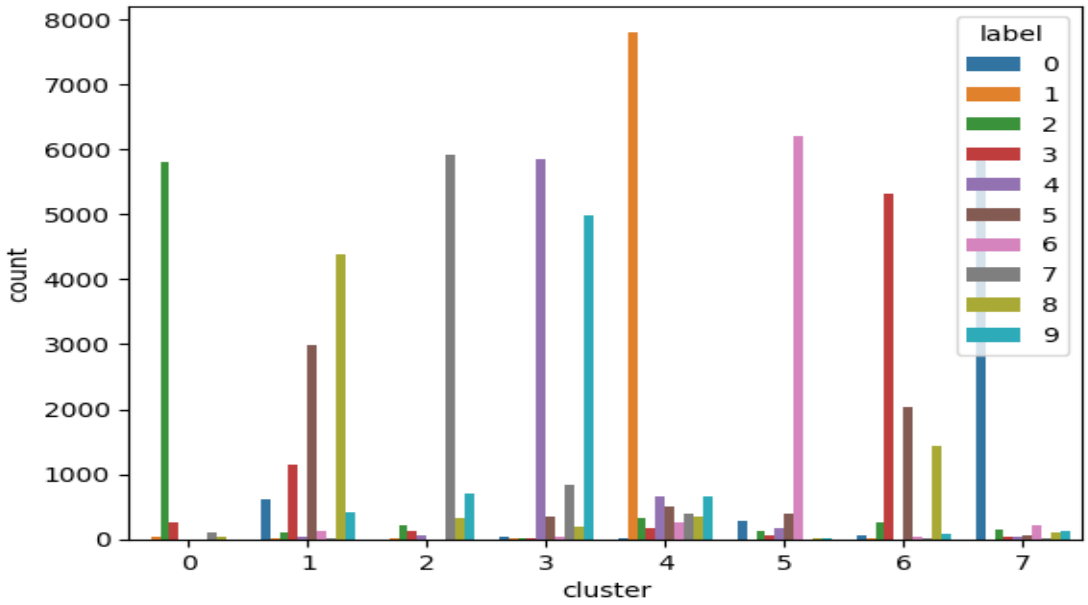
از تصویر 9، بهترین تعداد cluster برای داده های mnist، 8 است.



تصویر10. نمودار silhouette score براساس تعداد cluster برای دیتاست fmnist

از تصویر 10، بهترین تعداد cluster برای داده های fmnist، 8 است.

با تعداد بهینه کلاسترها، باردیگر الگوریتم Kmeans را اجرا می کنیم.

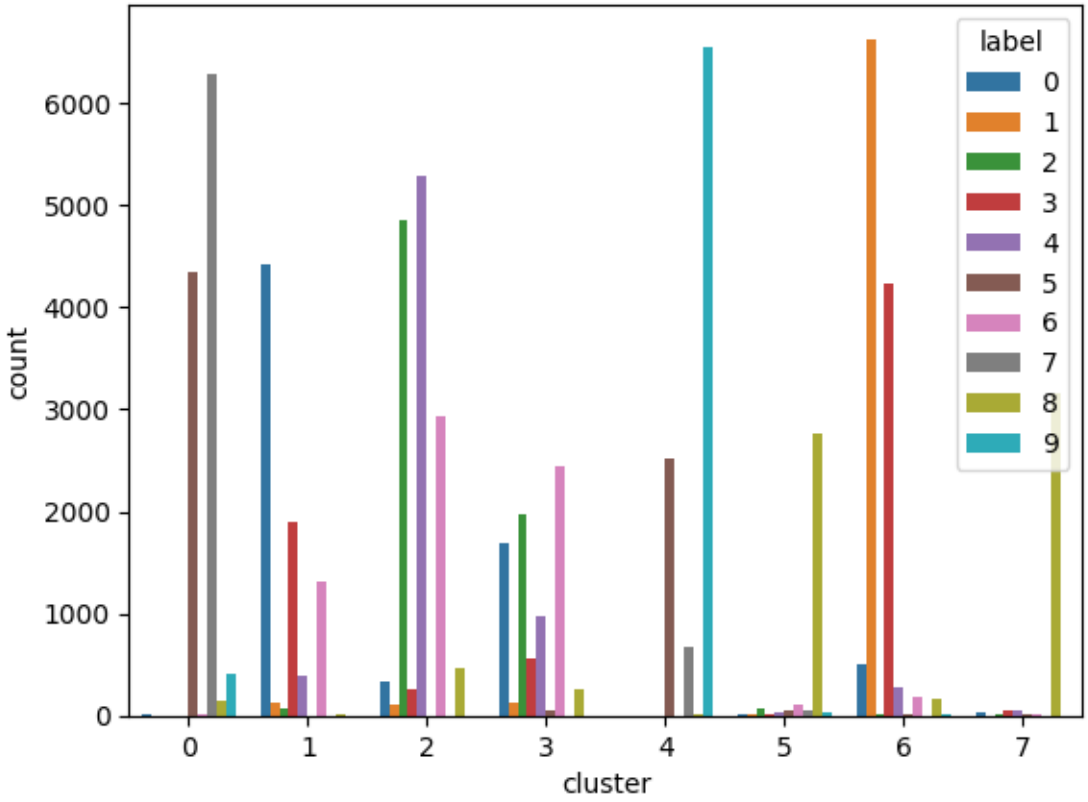


تصویر 11. ارتباط کلاسترها و لیبل تصویر برای داده mnist

جدول2. ارتباط کلاسترها و لیبل تصاویر برای داده mnist

کلاستر / لیبل	0	1	2	3	4	5	6	7
0								■
1					■			
2	■							
3							■	
4				■				
5		■						
6						■		
7			■					
8		■						
9				■				

طبق تصویر 11 و جدول 2، کلاستر 1 اغلب دارای لیبل های 5 و 8، و کلاستر 3 دارای لیبل های 4 و 9 است. علت این امر، شباهت بین شمایل 8 و 5، و 4 و 9، و یک دسته شدن توسط شبکه است. در دیگر کلاستر ها، لیبل غالب، یک لیبل معین است و شبکه به خوبی عمل جداسازی را انجام داده است. از طرفی، هر لیبل به صورت غالب، توسط یک و فقط یک کلاستر دسته بندی شده است.



تصویر12. ارتباط کلاسترها و لیبل تصویر برای داده fmnist

جدول3. ارتباط کلاسترها و لیبل تصاویر برای داده fmnist

کلاستر / لیبل	0	1	2	3	4	5	6	7
0		■		■				
1							■	
2			■	■				
3							■	
4			■					
5	■				■			
6			■	■				
7	■							
8						■		■
9					■			

داده های کلاستر 0، اغلب از کلاس 5 و 7 هستند. هردو این داده ها پاپوش (صندل و کتانی) هستند و به این علت در یک کلاستر قرار دارند. در کلاستر دو، اغلب داده های پلیور، کت و پیراهن قرار دارد که از لحاظ ظاهر و کاربرد یکسان اند. در کلاستر سه، اغلب داده های تاپ (تیشرت)، پلیور و پیراهن قرار دارند که همچنان ظاهر و کاربرد تقریبا یکسان دارند. در کلاستر چهار، اغلب داده های صندل و بوت قرار دارند که ظاهر تقریبا یکسان دارند. در کلاستر 6، اغلب داده های شلوار و دامن قرار دارد که از لحاظ اندازه و کشیدگی ظاهر تقریبا مشابه دارند. بقیه کلاسترها نیز دارای یک داده از دسته معین هستند. همچنین هر کلاس به صورت غالب توسط حداقل یک کلاستر تشخیص داده شده است.

همانطور که بررسی شد، هر دو شبکه روی هردو دیتاست، دسته بندی تقریبا مناسبی بر اساس شباهت های ظاهری داده ها انجام داده اند. احتمالا شبکه پیچیده تر با تعداد داده های بیشتر ورودی (یا آگمنتیشن) می توانست دقت بالاتری در دسته بندی داشته باشد.

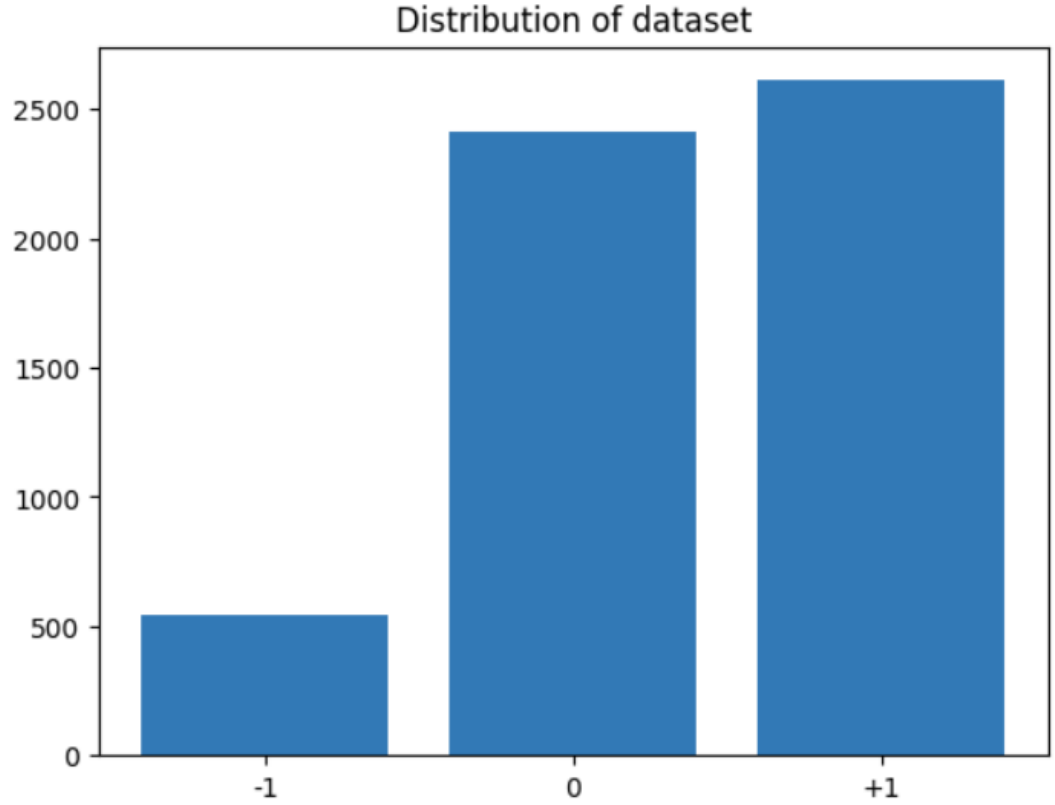
پرسش ۲ – افزایش داده در مدل FaBert

۲-۱. Data Augmentation در NLP

زمانی که داده‌های ما به اندازه کافی نمی‌باشد ما سعی می‌کنیم که با استفاده از روش‌هایی داده‌ها تقویب و افزایش بدیم تا مدل بهتر آموزش ببیند و دچار بیش برازش نشود که یکی از این روش مثلا می‌تواند قرار دادن کلمات مترادف به جای کلمات اصلی و یا جابه‌جایی تصادفی کلمات و یا حذف تصادفی کلمات در جمله باشد و یکی از این روش‌ها Back translation است که در این روش متن را به یک زبان دیگر ترجمه و سپس به زبان اولیه ترجمه می‌کنیم که باعث تغییر ساختار جملات و همچنین طرز بیان جدید می‌شود و می‌تواند داده‌های ما را افزایش دهد

۲-۲. پیش پردازش دادگان

دیتای ما مربوط به کامنت‌های دیجی‌کالا می‌باشند که در حالت مثبت و منفی می‌باشند(البته بعضی مثبت تر و بعضی منفی تر اما ما سه دسته خنثی،مثبت،منفی در نظر می‌گیریم)

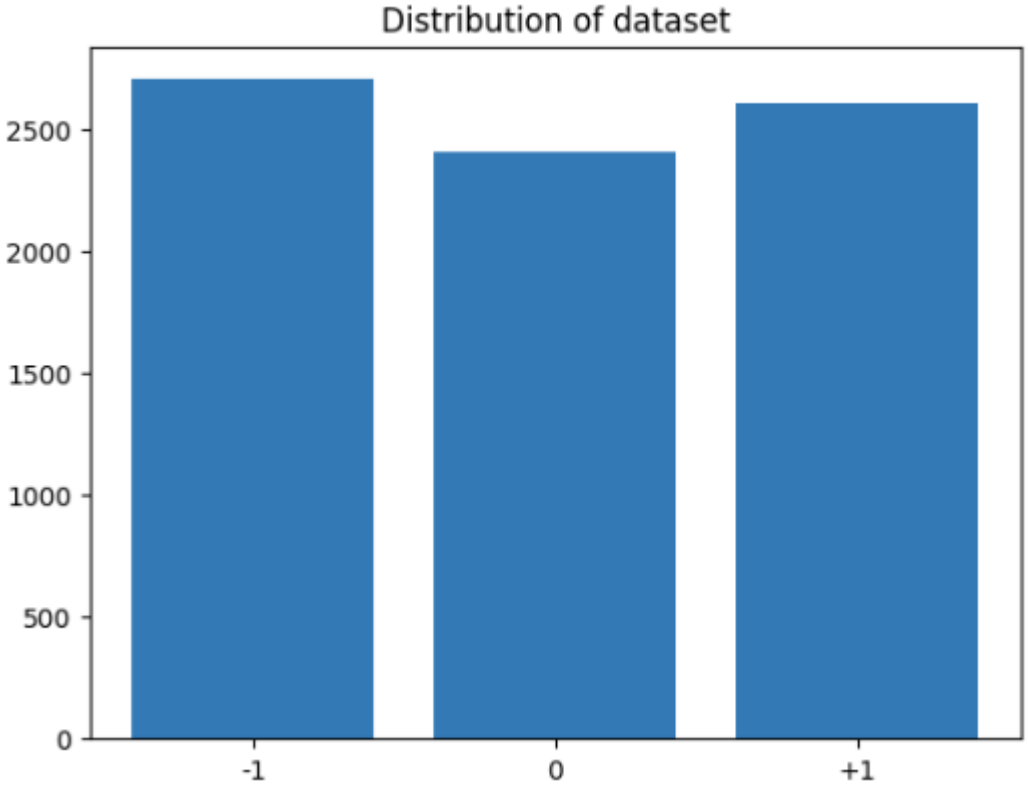


تصویر ۱. توزیع آماری کلاس‌های مجموعه داده

```
5560
count    5560.000000
mean       22.634532
std        20.403192
min         1.000000
25%        11.000000
50%        19.000000
75%        28.000000
max       320.000000
Name: num_tokens, dtype: float64
```

تصویر ۲. تعداد توکن‌ها

که مشاهده می‌شود داده‌های ما نامتوازن می‌باشند و دسته بندی کلاس‌های یکنواخت نیست همچنین میانگین توکن‌ها ۲۲ تا می‌باشد
از آنجایی که پس از ران بر روی همین دیتا ست مشاهده شد که به علت نامتوازن بودن داده‌ها و کم بودن نسبی داده‌های کلاس منفی یک آنگاه مدل ما اصلا برای کلاس منفی یک یادگیری ندارد و معیار f1 ما در حالت macro (که برای حالت نامتوازن معیار ارزیابی خوبی است چرا که برای کلاسای ارزش یکسان قائل می‌شه) نسبتا کم می‌شود پس تصمیم می‌گیریم که داده‌های کلاس منفی یک را زیاد کنیم به این منظور از آنجایی که تعداد اونها حدود ۵۰۰ و دو کلاس دیگر نزدیک ۲۵۰۰ پس داده‌های کلاس منفی یک را ۵ بار تکرار می‌کنیم تا به رنج دو کلاس دیگر برسد.



تصویر ۳. توزیع آماری کلاس‌های مجموعه داده پس از افزایش کلاس منفی یک

همچنین یک دهم داده‌های آموزشی را برای مجموعه ارزیابی در نظر می‌گیریم.

۳-۲. افزایش دادگان به روش Back translation

همانطور که گفتیم برای این امر ابتدا داده‌ها با استفاده از گوگل ترنسلیت به انگلیسی و سپس به فارسی برمی‌گردونیم

text:
در نهايت بايد بگم ديجي كالا كارش درسته و ميدونه چي معرفي كنه
back_translated_text:
در پایان باید بگویم دیجی کالا کار خود را به خوبی انجام می دهد و می داند چه چیزی را معرفی کند

text:
با برنامه‌های کاربردی برای تندرستی Fitness - سامسونگ می‌توانید برنامه‌های ورزشی و تمرینی خود را در منزل اجرا کنید و با قابلیت برنامه‌های کاربردی برای خانواده Family Story - سامسونگ می‌توانید مطمئن باشید که لحظات و خاطرات ارزشمند خانوادگی نه تنها ذخیره می‌شوند بلکه توسط تلویزیون هوشمند شما به اشتراک گذاشته خواهند شد.
back_translated_text:
با اپلیکیشن های سلامتی Samsung Fitness - می توانید برنامه های ورزشی و تمرینی خود را در منزل اجرا کنید و با قابلیت اپلیکیشن های خانوادگی - Samsung Family Storyمی توانید مطمئن باشید که لحظات و خاطرات ارزشمند خانوادگی نه تنها ذخیره می شوند، بلکه به اشتراک گذاشته می شوند. توسط تلویزیون هوشمند شما گذاشته خواهد شد

text:
اصلا فكر نكنين كه واسه بازي هاي آنلاين نياز به سرعت افسانه هاي دارين.
back_translated_text:
فکر نکنید که برای بازی های آنلاین به سرعت افسانه ای نیاز دارید.

text:
این گوشی مجهز به WLAN برای اتصال به شبکه از طریق سرویس‌های وایرلس محلی و GPS داخلی برای استفاده از انواع اپلیکیشن‌های GPS محور هم می‌باشد.
back_translated_text:
این گوشی مجهز به WLAN برای اتصال به شبکه از طریق سرویس های بی سیم محلی و GPS داخلی برای استفاده از انواع برنامه های GPS محور است.

text:
تجربه ي شخصي ام نشان مي دهد كه اگر چه گالكسي اس تري گوشي محبوب و خوبي هست ولي آيفون واقعا چيز ديگري است و از لحاظ سهولت كاربري نرم افزار iosو گستره بي نظير نرم افزارهاي مرتبط و امكاناتي مانند پيام رساني ، سرچ در دفترچه تلفن و كليه محتويات و نوت هاي ايجاد شده و كتب الكترونيكي ، سهولت تماس، قابليت بسيار آسان در وبگردي و كپي برداري از صفحات وب ، آيفون برنده ي بي رقيب ميدان است.
back_translated_text:
تجربه شخصی من نشان می دهد که اگرچه گلکسی اس گوشی محبوب و خوبی است، اما آیفون واقعا چیز دیگری است و از نظر سهولت استفاده از نرم افزار iOSو گستره بی نظیر نرم افزارها و امکانات مرتبط مانند پیام رسانی، جستجو در دفترچه تلفن و تمامی مطالب و یادداشت های ایجاد شده و کتاب های الکترونیکی، سهولت در تماس، قابلیت بسیار آسان در وب گردی و کپی صفحات وب، آیفون برنده بی رقیب این میدان است.

text:
ولي مطمئن شدم كه ion بهتر از acro s هست.
back_translated_text:
اما مطمئن شدم که یون بهتر از acro s است.

text:
"محدوديت بزرگترين كابوس بشر است!
back_translated_text:

"محدودیت بزرگترین کابوس بشریت است!

text:
من بین Nikon Coolpix L870 و این گیر افتادم اما دارم پولامو جمع و جور میکنم که این رو بخرم فکر میکنم گزینه بهتریه

back_translated_text:
من بین نیکون Coolpix L870 و این یکی دویده ام، اما برای خرید این یکی پس انداز می کنم، به نظرم گزینه بهتری است.

text:
در مقابل رقبایی مانند iPad mini یا Nexus 7 و یا حتیFonePad ، از نظر کیفیت ساخت باید Note 8 را در آخرین رده قرار دهیم.
back_translated_text:
در مقابل رقبایی مانند iPad mini یا Nexus 7 یا حتی FonePad باید نوت 8 را از نظر کیفیت ساخت در رده آخر قرار دهیم.

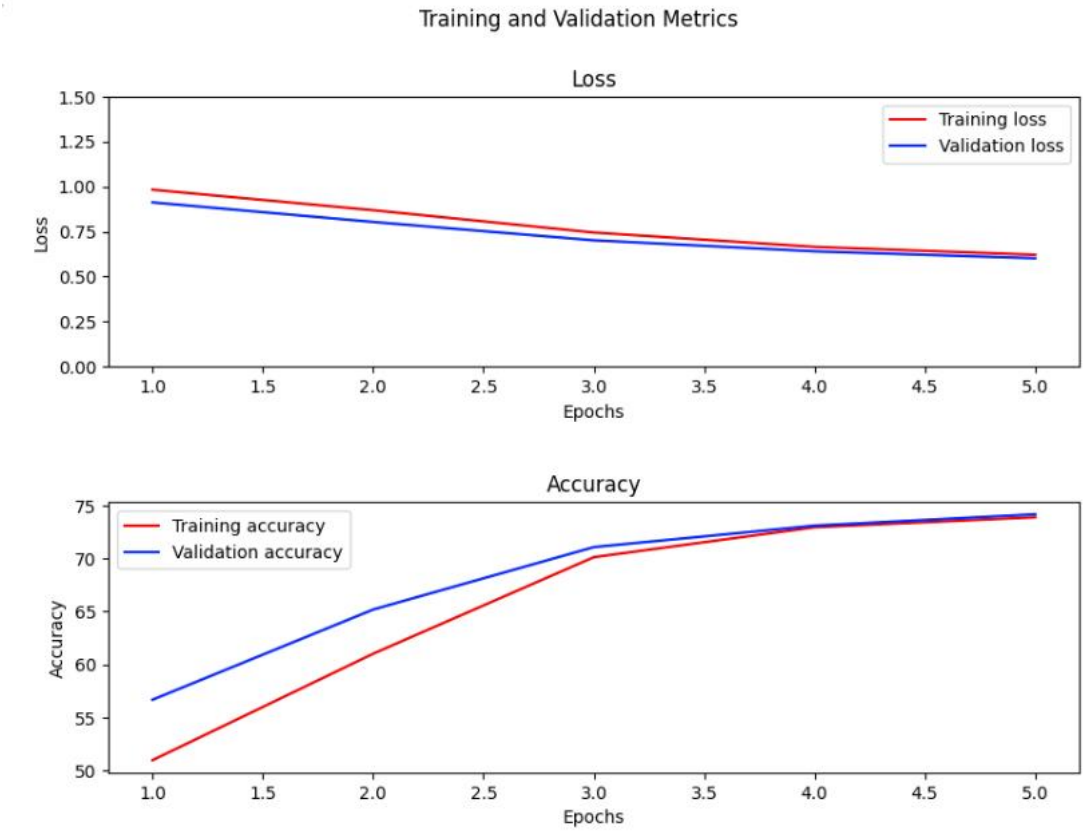
تعداد ۱۰ تا از جمله‌ها را قبل و بعد از این تغییر در بالا نمایش دادیم و بنظر می‌رسد که به خوبی وظیفه خود را انجام داده و کیفیت قابل قبولی دارند.

۲-۴. تنظیم دقیق(fine-tune) مدلFaBert

می‌خواهیم از یک مدل ترنسفور آموزش دیده بر روی دادگاه فارسی استفاده کنیم پس ابتدا ورودی‌ها را توکن می‌کنیم و از جنس عدد تا آماده ورود به مدل باشند و همچنین چون مسئله ما سه کلاسه است پس خروجی را بر این اساس می‌گذاریم(چون باید اعداد خروجی بین ۰ تا ۲ باشند برای سه کلاس پس کلاسای منفی یک را با ۲ نشان می‌دهیم) حال مطابق گفته‌های صورت سوال لایه کلاسیفیکشن و لایه آخر انکودر برت را آن فریز می‌کنیم که توانایی یادگرفتن داشته باشند و باقی لایه‌ها را بر اینکه مدل بیش‌از حد پیچیده نشود و دچار بیش برازش نشویم فریز می‌کنیم و هایپر پارامترای نرخ یادگیری(بهینه‌ساز را adamW در نظر می‌گیرم)و تعداد ایپاک را مشابه گفته شده قرار می‌دهیم

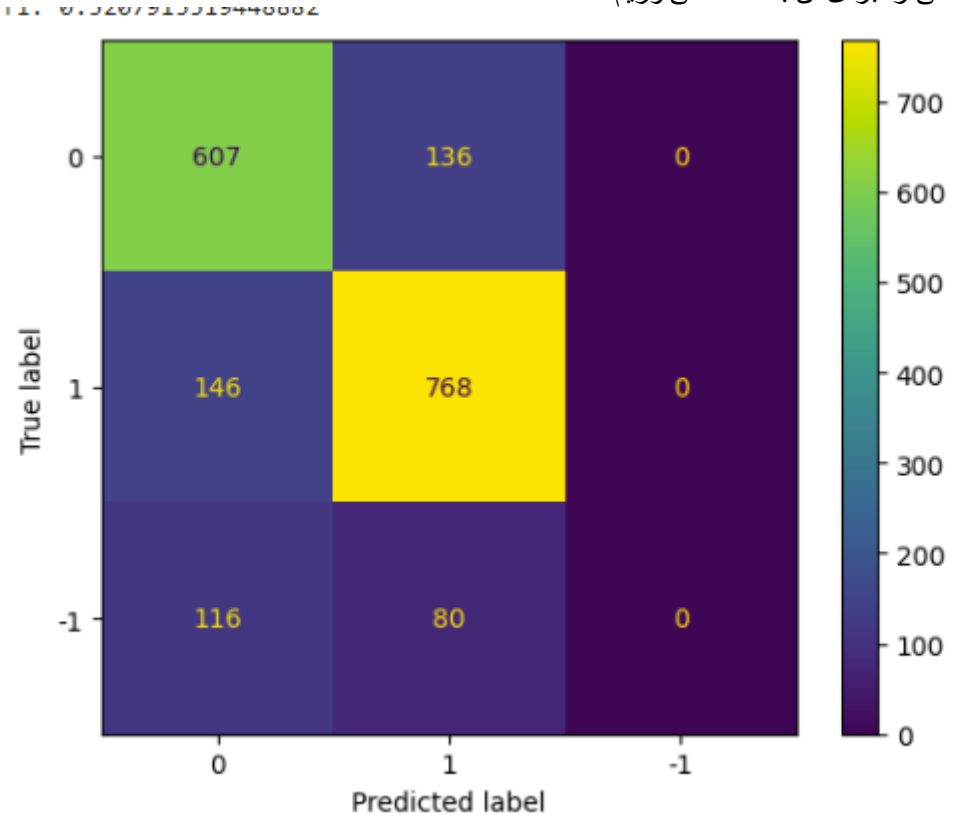
۲-۵. ارزیابی و تحلیل نتایج

ابتدا مدل را برروی دیتای اولیه و متوازن نشده ران می‌کنیم



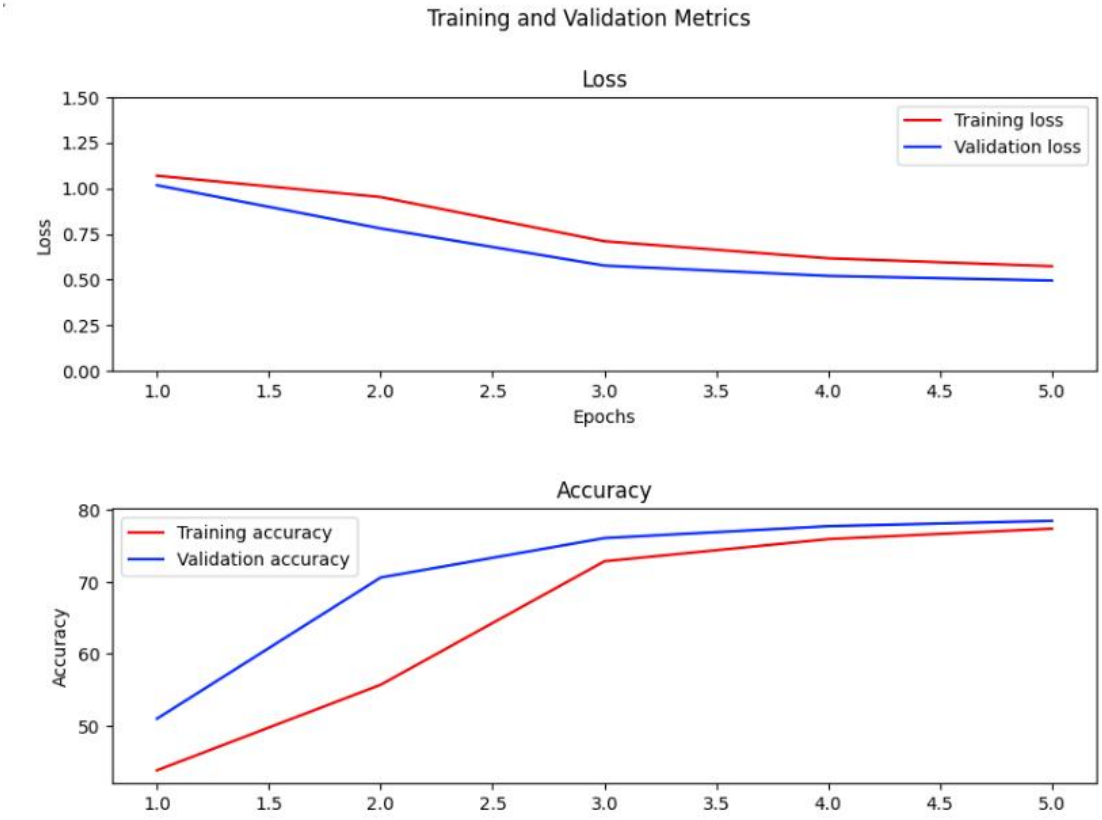
تصویر۴. نمودار تغییرات دقت و زیان بر روی دیتا اولیه شده

حال ماتریس آشفتگی را برای آن به دست می‌اوریم

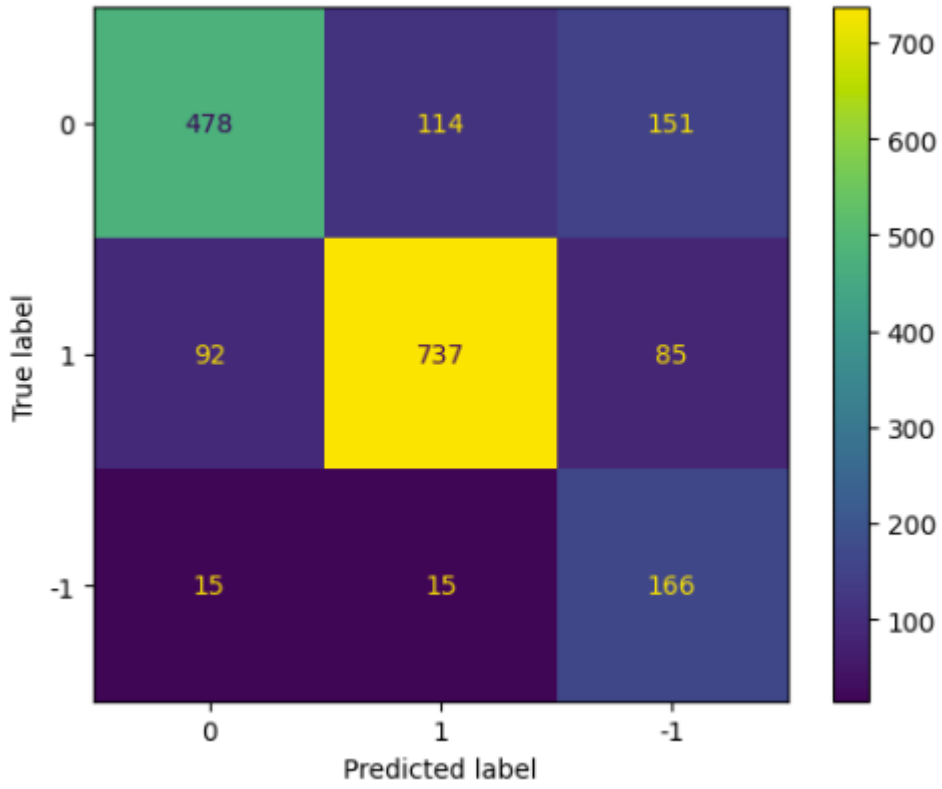


تصویر۵. ماتریس آشفتگی دیتاهای تست مدل آموزش دیده بر روی دیتای اولیه

همانظور که مشاهده می‌شود چون داده‌های ما نا متوازن بودن اصلا برای کلاس منفی یک عملکرد مطلوبی ندارد و کلا یاد نمی‌گیرد پس به درد نمی‌خورد با توجه به اینکه مخصوصا کامنتای منفی معمولا برای ما مهم تر است به همین علت داده‌ها را متوازن می‌کنیم
حال مدل را برروی داده‌های متوازن شده ترین می‌کنیم.

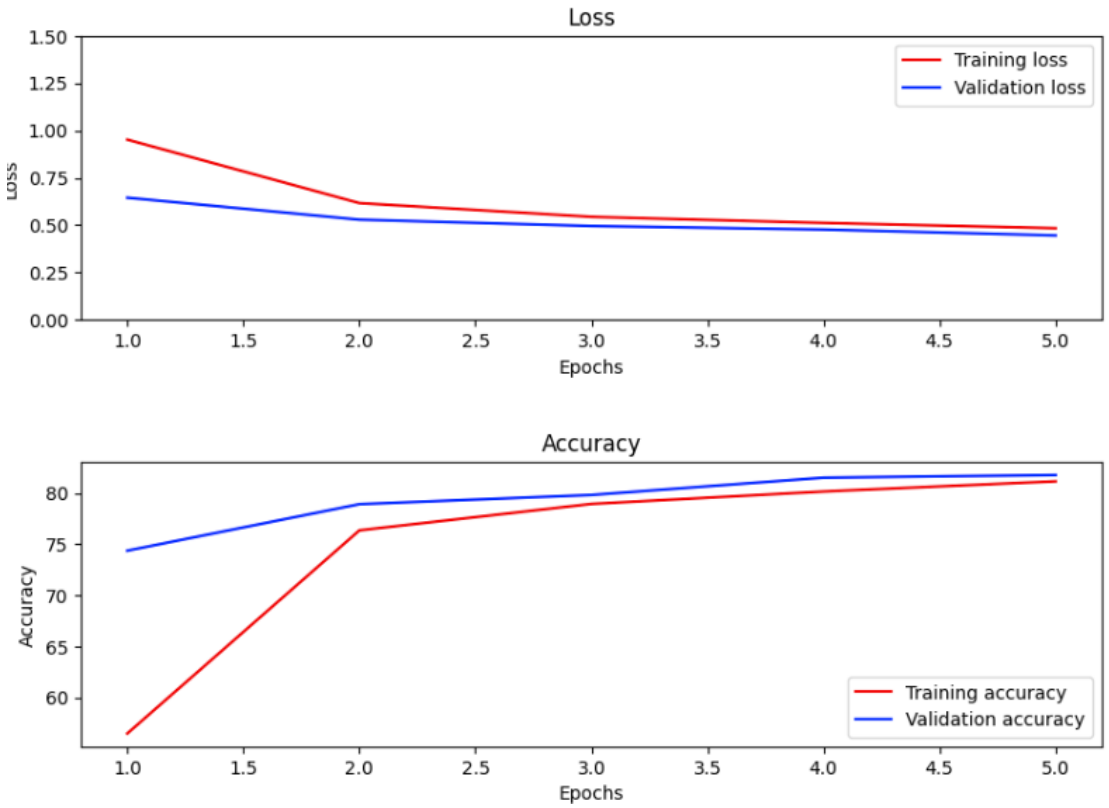


تصویر۶. نمودار تغییرات دقت و زیان بر روی دیتا متوازن شده

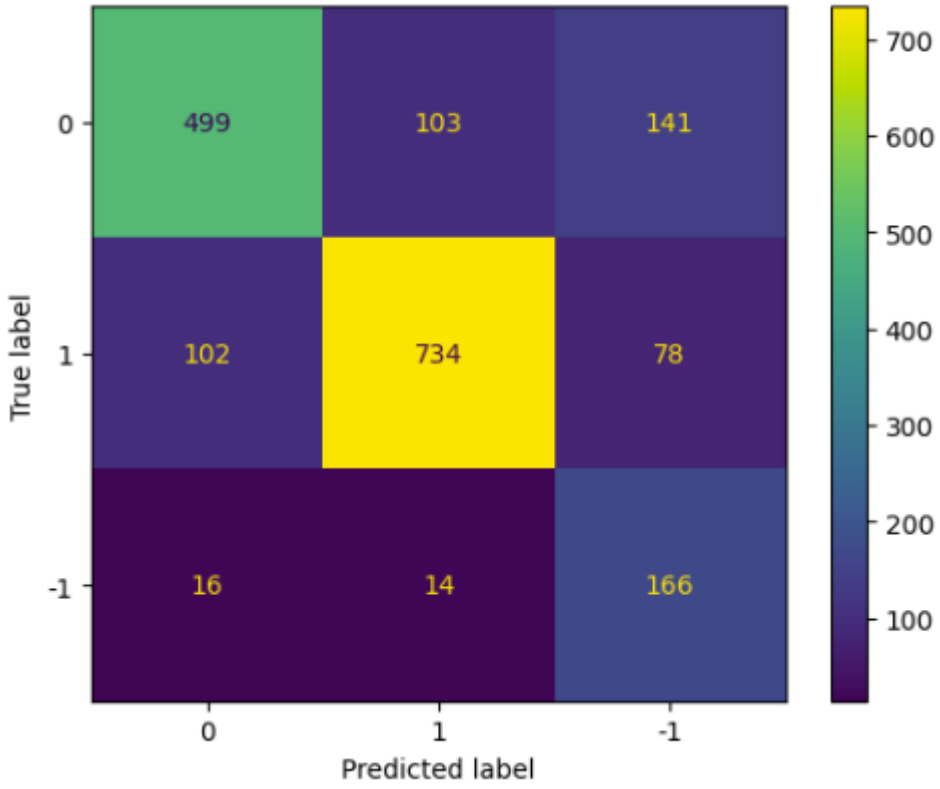


و همانطور که در ماتریس آشفتگی ملاحظه می‌شود مدل نسبت به حالت قبل بسیار بهتر عمل کرده و تا حدی به خوبی توانسته کلاس‌ها را تشخیص دهد و مشاهده می‌شود که معیار f1 ، ۷۵ شده است که نسبتا خوبه و همچنین دقت برروی داده‌های ارزیابی نیز بع ۷۳ رسیده است

حال به سراغ دیتای افزایش یافته می‌رویم که انتظار داریم در این حالت مدل ما بهتر عمل کرده باشد چراکه دیتای بیشتر و همینطور طرز بیانای بیشتری را دیده است



تصویر۸. نمودار تغییرات دقت و زیان بر روی دیتا افزایش یافته



تصویر۹. ماتریس آشفتگی دیتاهای تست مدل آموزش دیده بر روی دیتای افزایش یافته

در این حالت دقت برروی داده‌های ارزیابی به ۸۱ می‌رسد و همچنین معیار f1 به ۷۸ میرسد که ۵ درصد افزایش نسبت به حالت قبل را دارد و همانطور که انتظار داشتیم مدل ما بهتر آموخته شده است

حال تعدادی از داده‌های اشتباه پیش بینی شده را بررسی می‌کنیم

True Label: 2, Predicted Label: 0
البته يه بدي هم داره.

True Label: 1, Predicted Label: 0
ضبط ویدیویی هم می‌تواند با رزولوشن VGA و سرعت 25 فریم بر ثانیه انجام گیرد.

True Label: 1, Predicted Label: 0
اولا: تصوير برداري با ان بسيار سخت ميباشد به دليل نداشتن دكمه تله وايد روي دستگاه و سفت بودن رينگ زوم روي لنز اورجينال كه چرخاندن ان و تله وايد كردن باعث تكان خوردن دوربين شده و تصوير را خراب ميكند و در ضمن صداي ان هم درتصوير ضبط ميشود.

True Label: 0, Predicted Label: 1
سيستم عامل اندرويد خيلي باگ ها داره كه هنوز هم برطرف نشده ، براي مثال برنامه ي اتصال به اينترنتش يا برنامه ي مپ كه بسيار كند و كسل كننده است.

True Label: 1, Predicted Label: 2
من تا الان بیش از 5000 عکس باهاش گرفتم. من واقعا ازش راضیم و خیلی دوستش دارم. اگه باید عوضش کنم با a77 عوضش میکنم. با تشکر از دیجی کالا که در انتخاب کمک زیادی می کند.

True Label: 0, Predicted Label: 1
از صداش هم خوشم نیومد.

True Label: 0, Predicted Label: 1
این بار نیز این کمپانی طراحی و تولید تبلت خود را به شرکت Asus سپرد که خود در حال حاضر یکی از بزرگترین و پر طرفدارترین کمپانی‌های تولیدکننده تبلت و لپ تاپ است.

True Label: 0, Predicted Label: 2
گذشته از ظاهر شيک و سنگينش كيفيت بدنه موقعي كه اون رو در دست داريد احساس ميشه.

True Label: 1, Predicted Label: 2
سلام.

True Label: 1, Predicted Label: 2
یکی از قسمت‌هایی که در محصولات دیجیتالی، همواره از اهمیت ویژه ای برای ایسوس برخوردار بوده است، کیفیت صوتی دستگاه است.

True Label: 0, Predicted Label: 2
از خوبی‌های گوشی‌های ساخت ال جی می‌توان به وجود ویجت‌های متنوع و کاربردی موجود در گوشی‌های ساخت این کمپانی اشاره نمود.

همانطور که مشاهده می‌شود بنظر می‌رسد در دیتای اشتباه پیش بینی شده تا حدی کامنتای لیبل زده شده به حدی سختگیرانه می‌باشند و شاید خدود ماهم اشتباه می‌کردیم در مجموع بنظر می‌رسد که مدل ما به خوبی کار کرده است و تا حدی معقول تونسته است یادبگیرد و افزایش دادگان و متوازن کردن داده‌ها موجب بهبود مدل ما می‌شود

پرسش ۳ – کلمه‌ی بیدار باش

3-۱. جمع آوری داده

با استفاده از دو تابع record_wake_word و record_background_sound، به ترتیب 100داده صوتی با طول 2 ثانیه از کلمه "Awake" و 100 داده صوتی از کلمات دیگر و صدای "Background" ضبط می کنیم.
صداها در این لینک قرار دارند.

3-2. پیش پردازش و استخراج داده

پیش پردازش:

سیگنال صوت به صورت آنالوگ است. ابتدا برای پیش پردازش، با نمونه گیری از داده های آنالوگ، آن را به داده های دیجیتال تبدیل می کنیم. در این بخش از فرکانس 44.1 kHz. سپس داده های نمونه گیری شده را از بازه اعداد پیوسته به بازه اعداد گسسته (با تعیین فرمت 16بیتی یا 24بیتی) انتقال می دهیم.
داده های استخراج شده دارای سکوت هستند. سطح سکوت را در اینجا db20 تعیین می کنیم و Amplitude پایین تر از این مقدار را حذف می کنیم. داده های ورودی از Amplitudeهای متفاوت هستند. پس داده های ورودی را نرمالایز می کنیم.
داده های ورودی به صورت نمودار Amplitude برحسب زمان هستند. داده هارا با استفاده از windowing، سگمنت می کنیم. روی هر سگمنت Fourier Transform اعمال می کنیم تا به نمودار Spectrogram برسیم. به این عمل short term fourier transform گفته می شود. سپس Spectrogram را به melspectrogram تبدیل می کنیم تا به بازه فهم صوت انسان دست یابیم. مجدد داده های melspectrogram نرمالایز می کنیم.

ویژگی های صوت:

صوت دارای ویژگی های زمان محور و ویژگی های فرکانس محور است. ویژگی های زمان محور مثل Waveform (داده خام صوت که میزان amplitude بر اساس زمان است)، انرژی (تابعی از amplitude در یک بازه زمانی، که می تواند نمایانگر بلندی صدا یا وجود صدا در محیط، در یک بازه باشد)، Zero-Cross rating (فرکانسی که amplitude از خط معیار می گذرد، می تواند در تشخیص حروف صدادار یا در تسک تشخیص ژانر موزیک مفید باشند) هستند.
ویژگی های فرکانس محور مثل Spectrogram (که همانطور که بالاتر اشاره شد، حاصل STFT روی صوت، به صورت فرکانس بر حسب زمان است، و کاربرد زیادی در تسک های CNN-based، از جمله دسته بندی صوت دارد)، Mel-Frequency Cepstral Coefficients (که حاصل تبدیل Spectrogram به بازه فهم شنوایی انسان است و کاربرد مشابه Spectrogram دارد) هستند که حاصل پیش پردازش روی ویژگی های زمان محور (به صورت خاص waveform) است.
ویژگی های دیگر از جمله گشتاور های زمانی (که حاصل گشتاورهای احتمالی از جمله میانگین و انحراف از معیار روی waveform در بازه معین است و تحلیل آماری صوت را ممکن می کند)، Autocorrelation (که شباهت صوت در یک زمان با مقادیر همان صوت در زمان های قبلی است، که در تعیین فرکانس صوت کمک می کند)، Short-Term Energy (جمع انرژی در بازه زمانی معین است که می تواند سکوت، بلندی صدا و تغییر آن در طول زمان را تشخیص دهد)، به عنوان ورودی، اطلاعات زیادی باوجود ویژگی های spectrogram و Mel-Frequency Cepstral Coefficients اضافه نمی کنند و افزودن آن ها باعث redundancy در مدل می شود.

دیتا آگمنتیشن:

روش آگمنتیشن زمان محور:

در Time Streching، سرعت صوت را بدون تغییر فرکانس صدا، تغییر می دهیم.
در Timeshift Spectrogram، نمودار spectrogram را به صورت افقی جابجا می کنیم. با این کار، سعی می کنیم شبکه را نسبت به زمان بیان کلمات مقاوم کنیم.

روش آگمنتیشن فرکانس محور:

در Frequency Shift Spectrogram، نمودار spectrogram را به صورت عمودی جابجا می کنیم. با اینکار شبکه را نسبت به فرکانس های متفاوت از بیان کلمات مقاوم است.

روش آگمنتیشن نویزی:

در Add Noise، به صورت رندوم، به amplitude، مقادیری را اضافه یا کم می کنیم. با اینکار، شبکه نسبت به داده های دارای نویز در پس زمینه مقاوم می شود.

روش آگمنتیشن با ماسک کردن:
در Time Masking، قسمت هایی از محور زمان را در spectrogram ماسک می کنیم (صفر می کنیم).
در Frequency Masking، قسمت هایی از محور فرکانس را در spectrogram ماسک می کنیم.

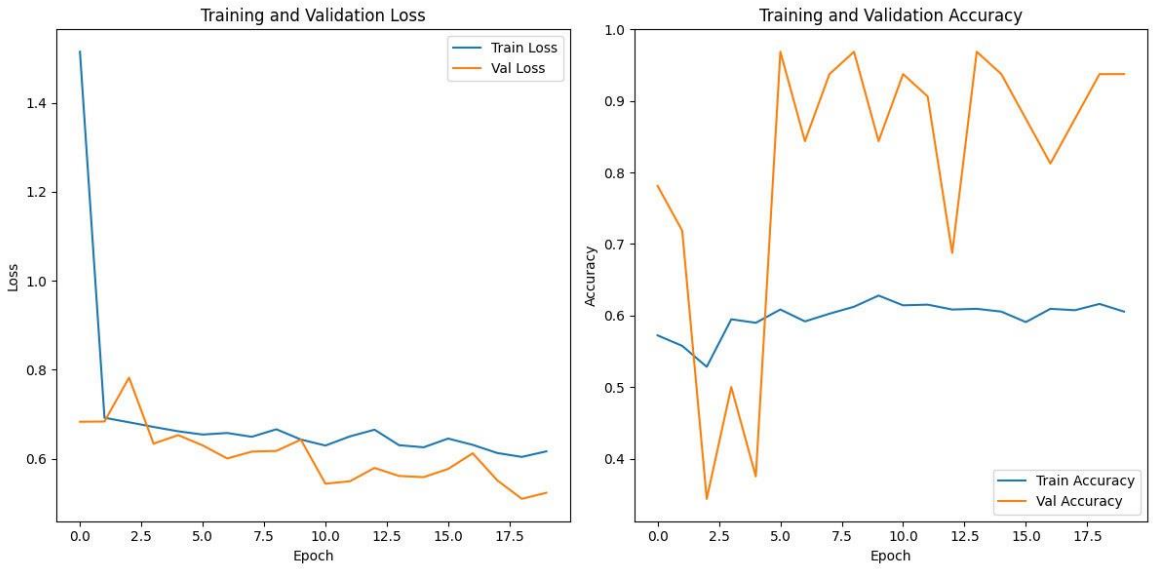
3-3. طراحی شبکه

اگر طول صوت یکسان باشد، می توانیم از RNNs یا از CNN استفاده کنیم. CNN در درک ویژگی های محلی از صوت موثر است و داده ورودی با ابعاد یکسان دریافت می کند. همچنین RNNs دردرک ویژگی های دنباله ای داده موثر است. بنابراین با وجود طول ثابت، می توان از هر دوشبکه استفاده کرد.
درصورت متغیر بودن طول صوت ورودی، صرفا از RNNs می توانیم استفاده کنیم. زیرا این شبکه ها قابلیت دریافت داده های ورودی با طول متغیر را دارند.
در این سوال طول داده ها ثابت است و با توجه به کوتاه بودن طول داده (2ثانیه) و توانایی CNN در درک ویژگی های محلی، شبکه CNN می تواند مفید باشد و کلمه "Awake" را تعداد لایه کم از کلمات دیگر تشخیص دهد.
تعداد 200 داده صوت (100 داده از کلاس "Awake" و 100 داده از اصوات دیگر) را به صورتی تقسیم می کنیم که 40 داده به عنوان داده test و 40 داده به عنوان داده validation و 120 داده به عنوان داده train قرار گیرند.
برای آموزش از 20 epochs و 8 batch size استفاده می کنیم. همچنین در نهایت بهترین مدل را براساس دقت روی داده های validation انتخاب می کنیم. از روش بهینه سازی adam و تابع هزینه binary-crossentropy استفاده می کنیم.

Layer (type)	Output Shape	Param #
conv2d_18 (Conv2D)	(None, 126, 85, 32)	320
dropout_12 (Dropout)	(None, 126, 85, 32)	0
max_pooling2d_18 (MaxPooling2D)	(None, 63, 42, 32)	0
batch_normalization_24 (BatchNormalization)	(None, 63, 42, 32)	128
conv2d_19 (Conv2D)	(None, 61, 40, 64)	18496
max_pooling2d_19 (MaxPooling2D)	(None, 30, 20, 64)	0
batch_normalization_25 (BatchNormalization)	(None, 30, 20, 64)	256
conv2d_20 (Conv2D)	(None, 28, 18, 128)	73856
max_pooling2d_20 (MaxPooling2D)	(None, 14, 9, 128)	0
batch_normalization_26 (BatchNormalization)	(None, 14, 9, 128)	512
flatten_6 (Flatten)	(None, 16128)	0
dense_12 (Dense)	(None, 128)	2064512
dropout_13 (Dropout)	(None, 128)	0
dense_13 (Dense)	(None, 1)	129

Total params: 2158209 (8.23 MB)
Trainable params: 2157761 (8.23 MB)
Non-trainable params: 448 (1.75 KB)

تصویر .شبکه CNN مورد استفاده



تصویر . نمودار دقت و خطا روی داده های train و validation

جدول .نتایج روی داده های تست

accuracy	loss
95%	0.60

جدول .ماتریس درهمریختگی برای داده های تست

1 (بیدار باش)	0	پیش بینی / کلاس
0	18	0
20	2	1 (بیدار باش)

.

پرسش ۴ – شبکه بخش بندی تصاویر

۴-۱. دادگان .

هدف ما در این تمرین segment کردن عکسا می‌باشد و به همین منظور دیتا ست ما که SUIM باشد شامل عکس و مسکای متناظر با آن‌ها می‌باشد که عکسای ما تصاویری از زیر آب می‌باشند و همچنین ۸ تا کلاس داریم که شی‌های ما در ۸ حالت می‌توانند باشند که مسکای ما می‌شوند از آنجایی که عکسای ما RGB هستند پس دیتا ست آمده و بر اساس اینکه این سه بعدی به یکی از ۸ حالت ۰۰۰ و ۰۰۱ و ۰۱۰ و ۰۱۱ و ۱۰۰ و ۱۰۱ و ۱۱۰ و ۱۱۱ است

حال در ابتدا برای یکسانی سازی و همچنین سبک شدن محاسبات سایز عکسا را تغییر می‌دهیم و همه را ۱۲۸ در ۱۲۸ می‌کنیم و همچنین به شکل زیر و با استفاده از کتابخونه albumentations به منظور جلوگیری از بیش برازش داده‌ها را تقویت می‌کنیم

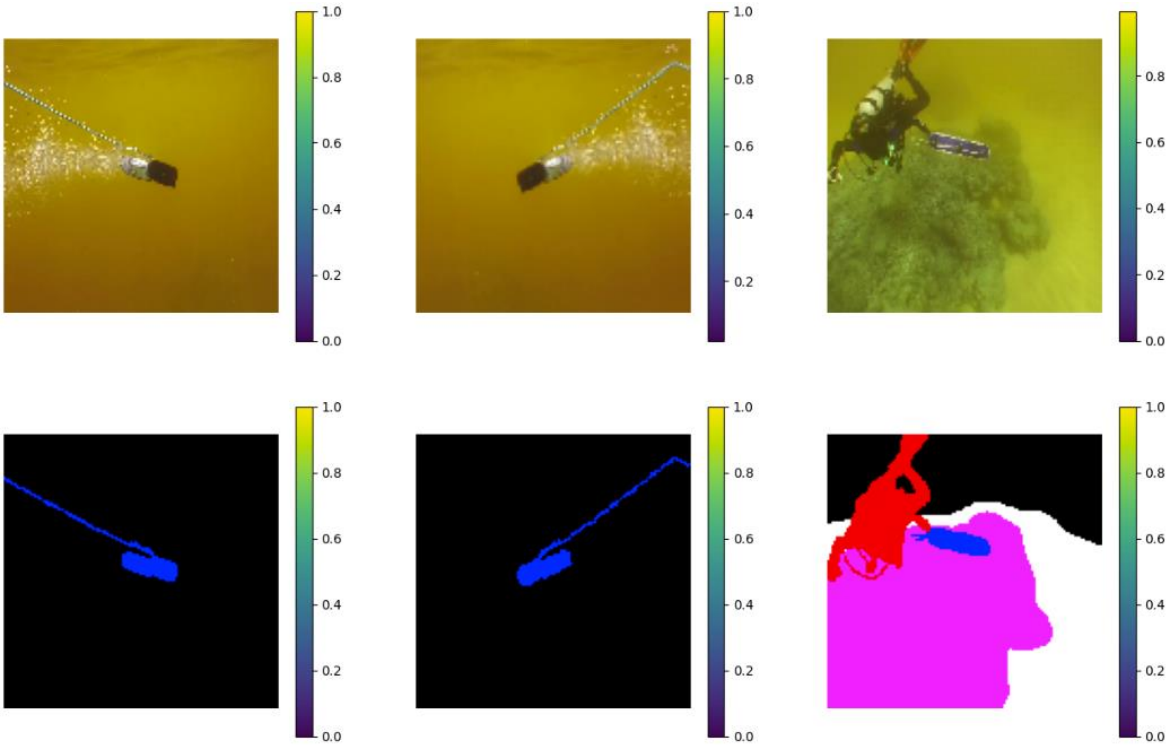
```
transform = A.Compose([
    A.HorizontalFlip(p=0.5),
    A.RandomBrightnessContrast(p=0.2),
    A.Rotate(limit=15, p=0.5),
    A.ShiftScaleRotate(shift_limit=0.1, scale_limit=0.1, rotate_limit=15, p=0.5)
])
```

شکل ۱. روش‌های تقویت داده

.حال ۱۰ درصد داده‌های آموزشی را به داده‌های اعتباری سازی می‌دهیم و ۹۰ درصد باقی مانده را برای داده‌های آموزشی نگه می‌داریم(و علت اینکه می‌خواهیم داده/ ها متوازن تقسیم شوند این است که زمانی که ترین می‌شوند داده‌های اعتبار سنجی معیار منصفانه‌ای باشند و همچنین مدل ما همه داده‌ها را بر اساس اهمیتشان دارد)

حال از آنجایی که گفتیم دوست داریم پیکسل‌های مسک‌های ما به شکل اون ۸ حالت باشند و بعداز این تغییرات پیکسل‌ها لزوما سه بعدشون ۰ و ۲۵۵ نیست حال پس یک مرز(ترشولد) برای پیکسل‌ها تعیین می‌کنیم و اگر پیکسلی در بعدی بیشتر از ۷۰ داشت آن را ۲۵۵ می‌کنیم.

و در نهایت تمامی تصاویر را با روش min-max normalization نرمال می‌کنیم که در واقع معادل این است بر ۲۵۵ تقسیم می‌کنیم تا اعداد ما بین ۰ تا ۱ باشند تا الگویتم گرادیان ما زودتر همگرا شود و اعداد بزرگ نشوند.

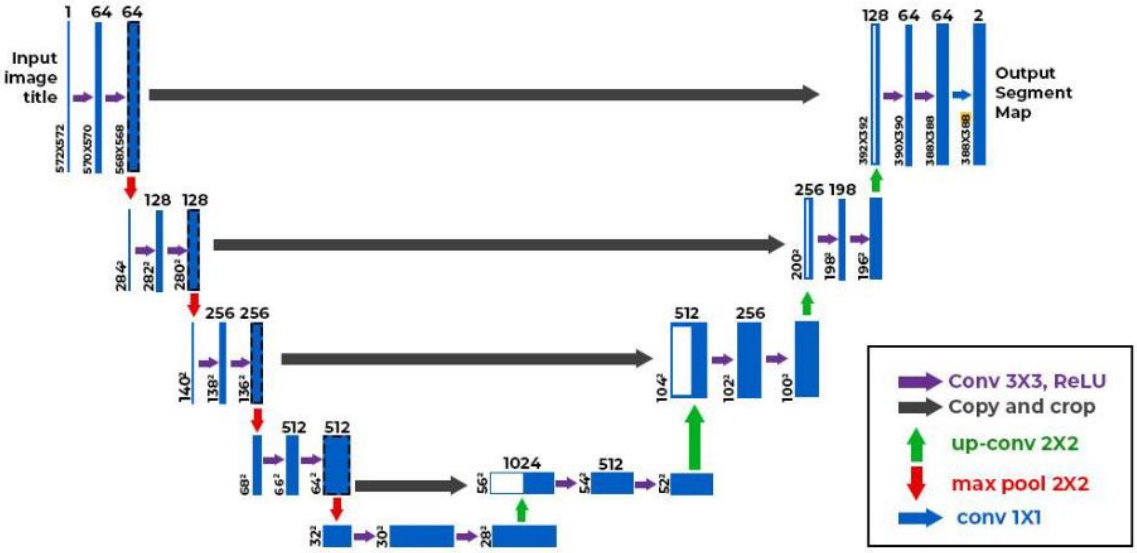


شکل ۲. نمونه‌ای از تصاویر و مسک آن‌ها

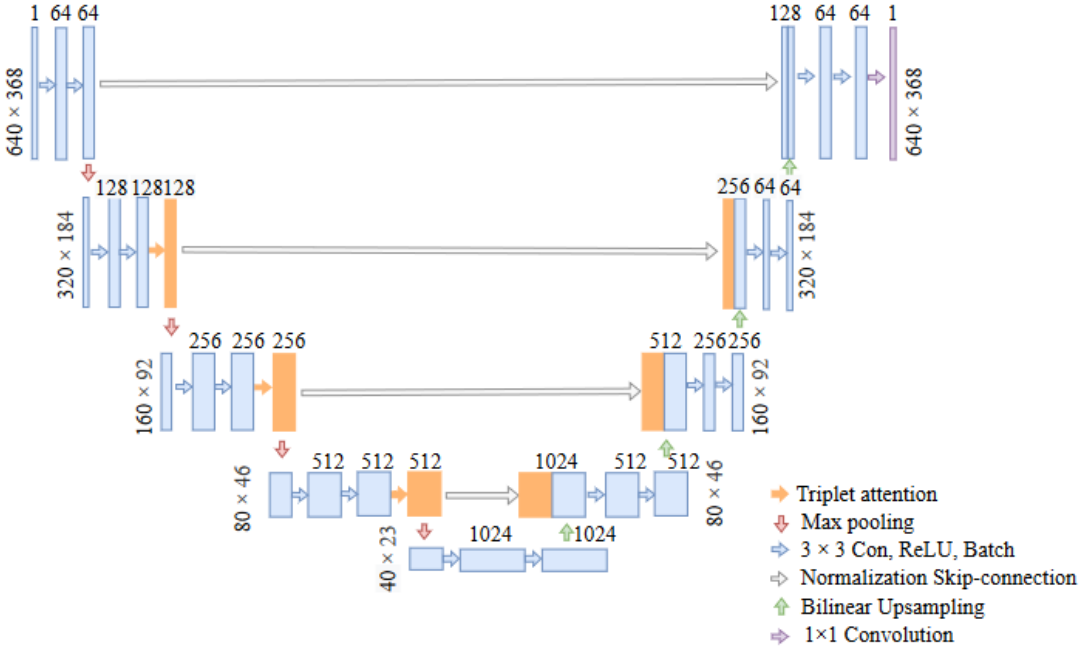
حال می‌دانیم که ۸ تا کلاس داریم و برای یادگیری چند کلاسه دوست داریم که از تابع softmax استفاده کنیم برای همین مسئله را اینطوری در نظر می‌گیریم که هر پیکسل می‌تواند یکی از ۸ کلاس باشد پس برای همین ابتدا عدد باینری حاصل از ۳ بعد تصویر مسک را به مبنای ۱۰ میآریم که عددی بین ۰ تا ۷ خواهد بود و سپس بر اساس آن بردار را به شکل وان هات در می‌آوریم که بتوانیم با تابع سافت مکس آن را نشان دهیم.

۴-۲. شبکه مورد استفاده

UNetیک شبکه عصبی کانولوشن است که برای مسائل segmentation تصویر، به خصوص در حوزه پردازش تصاویر ، طراحی شده است. از دو بخش اصلی انکودر و دیکودر تشکیل شده است که در انکودر پیژگی‌های تصویر یافته می‌شود و سپس در انکودر دوباره تصویر را بزرگ می‌کنیم و از روی ویژگی‌هایش آن را می‌سازیم TA_UNet از روی UNet گسترش سافته است و مفهوم attention را به آن اضافه می‌کند و اطلاعات مکانی را اضافه می‌کند و بین هردو بلوک انکودر و دیکودر قرار می‌گیرد و باعث می شه به اطلاعات مکانی و زمینه‌ای توجه بیشتری شود و به بخش‌های مهم تصویر توجه بیشتری شود.



شکل۳. ساختار شبکه Unet



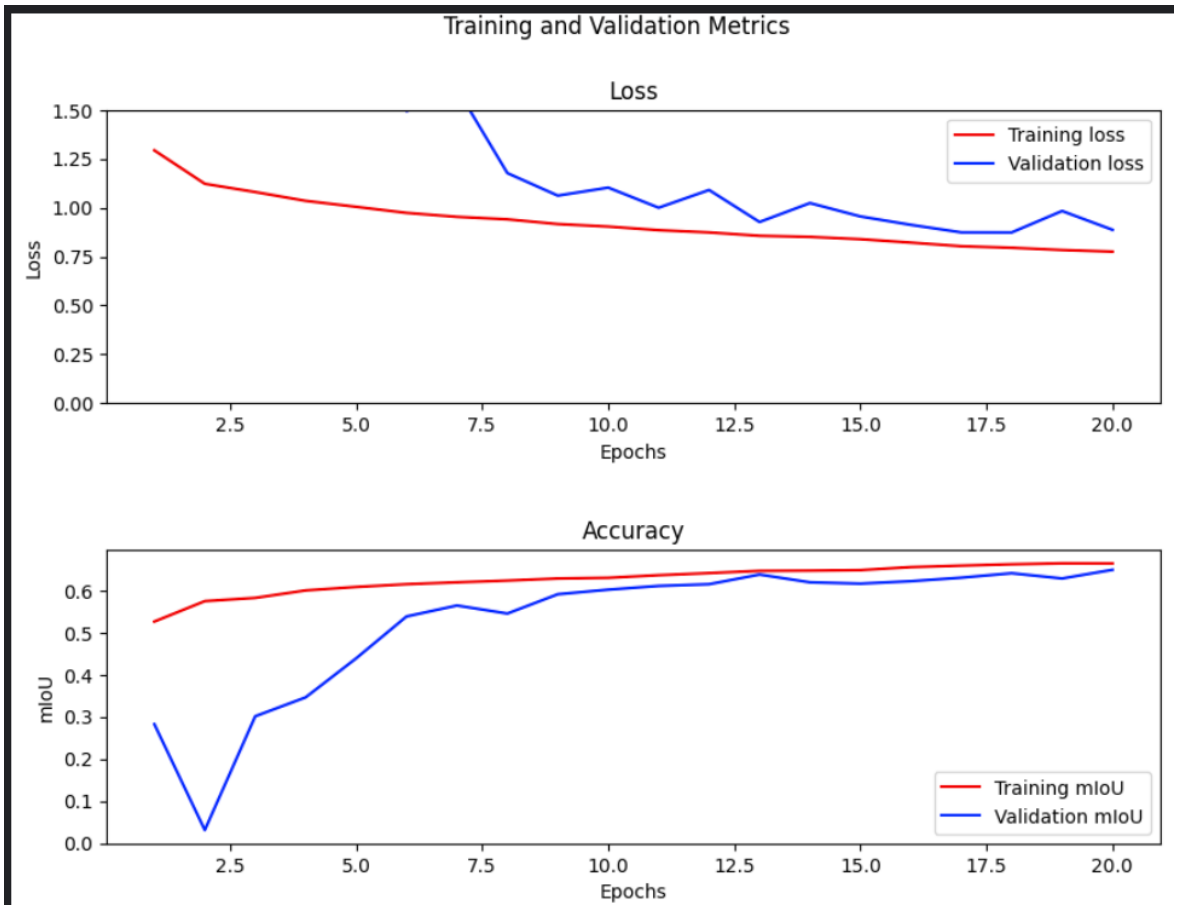
شکل۴. ساختار شبکه TA_Unet

۴-۳. آموزش شبکه

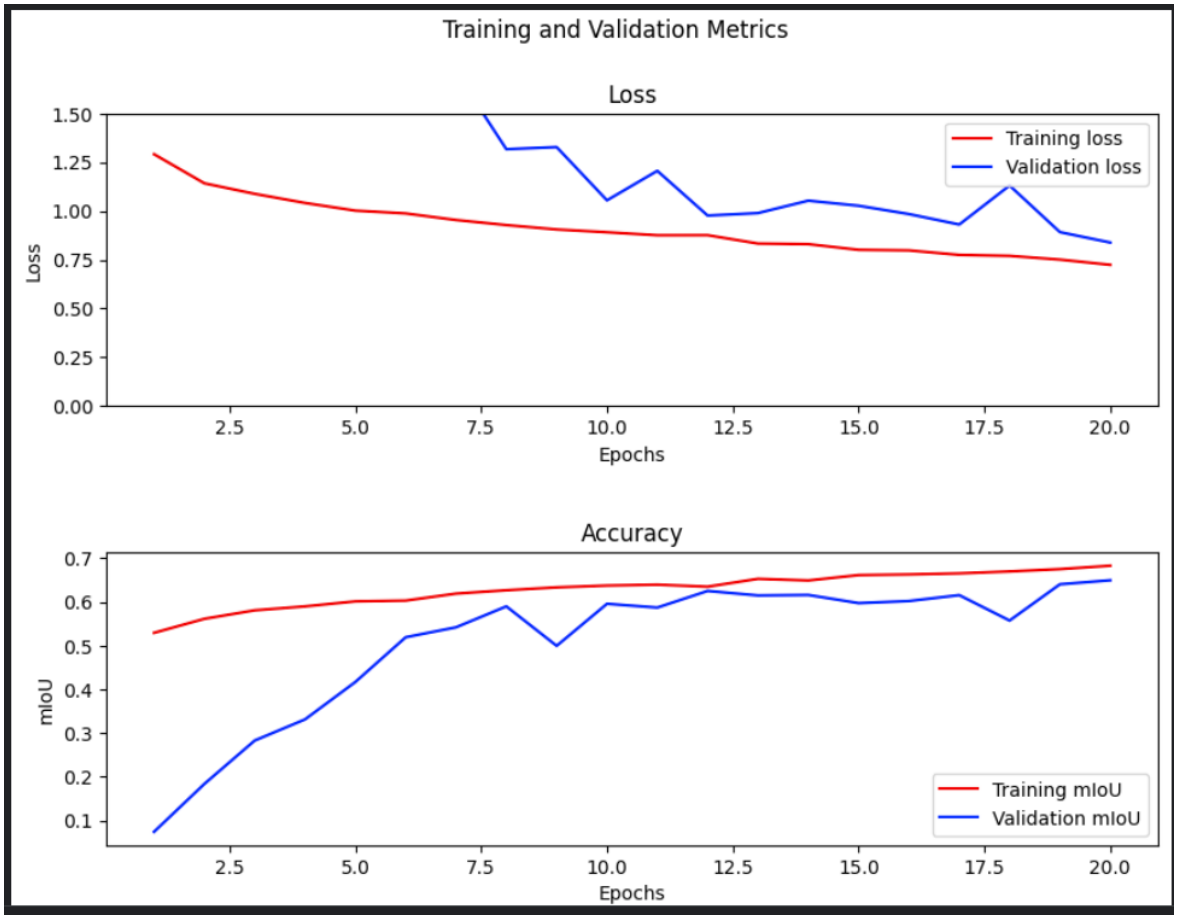
ساختار unet و tauent را مشابه مقاله پیاده سازی می‌کنیم و باهمان سایزها و همانطور که گفتیم جون تابع خروجی را سافت مکس در نظر گرفتیم پس از لاس کراس انتروپی استفاده می‌کنیم و همچنین بهینه‌ساز آدام و برای ۲۰ ایپاک می‌گذاریم ران شود

۴-۴. ارزیابی و تحلیل نتایج

در حالتی که در بخش بندی شامل چند کلاس هستیم از miou استفاده می‌کنیم که درواقع میانگین معیار Iou برای کلاس‌های مختلف می‌باشد. کهiou نشان دهنده درصد پیکسل‌هایی از کلاس که به درستی تشخیص داده شده‌اند بر تعداد کل پیکسل‌های درست این کلاس به علاوه پیکسل‌هایی که به اشتباه برای این کلاس معرفی شده‌اند.



شکل5. نمودار تغییرات زیان و miou برای unet



شکل6. نمودار تغییرات زیان و miou برای TA_unet

و مقدار miou برروی داده‌های تست برای هدو شبکه به ترتیب ۵۶ و ۵۸ درصد می‌شود

بنظر می‌رسد با اضافه کردن توجه و اهمیت بیشتری به مکان‌ها نتایج بهتری گرفتیم و TA_unet بکه بهتری است