

گزارش پروژه نهایی

درس یادگیری ماشین

رضا میرزایی ۹۶۲۱۰۰۵۹
دانیال علی حسینی ۹۶۲۰۱۴۴۶
دکتر بیگی
دی ۹۶

مقدمه

مساله ای که با آن رو به رو هستیم تشخیص و دسته بندی اعداد دست نویس مجموعه داده MNIST می باشد. می دانیم که بالاترین دقت در یادگیری با نظارت و بدون نظارت در این مجموعه داده با شبکه های عصبی بدست آورده شده است که چه به عنوان عنصر استخراج ویژگی با شبکه های عصبی کانوولوشنال چه برای دسته بندی شبکه های عصبی بهترین نتایج را داشته اند.

در این پروژه کوشیدیم فارغ از شبکه های عصبی و با دسته بندی هایی که در درس یادگیری ماشین آموختیم و پیاده کردیم به نتایج قابل قبولی از درصد درستی در یادگیری با نظارت و بدون نظارت دست یابیم و سعی کردیم ویژگی هایی از تصاویر در هر کدام از بخش ها استفاده کنیم که بهترین نتایج را به دنبال داشته باشند.

در قسمت یادگیری با نظارت از ویژگی SURF و دسته بندی های چون KNN و Random forest استفاده کردیم و همچنین از روش پنجره گذاری در تصویر برای مقایسه بهتر ویژگی ها استفاده شد که بسیار در بالا بردن دقت موثر بود و برای کاهش ابعاد هم از pca استفاده شد.

در قسمت یادگیری بدون نظارت از ویژگی HOG و همچنین خوشه بندی GMM استفاده شده است و با تغییر اندازه و سطح HOG و پارامتر های KNN و همچنین استفاده از توانستیم به دقت خوبی برسیم.

در ادامه گزارش شرحی از هر کدام از ویژگی های SURF و HOG ارائه می دهیم ، توضیح در مورد روش دسته بندی و خوشه بندی مورد استفاده می نگاریم ، پنجره گذاری در تصویر و منطق آن را توضیح می دهیم و در پایان نتیجه های که دست یافتیم را به اشتراک می گذاریم.

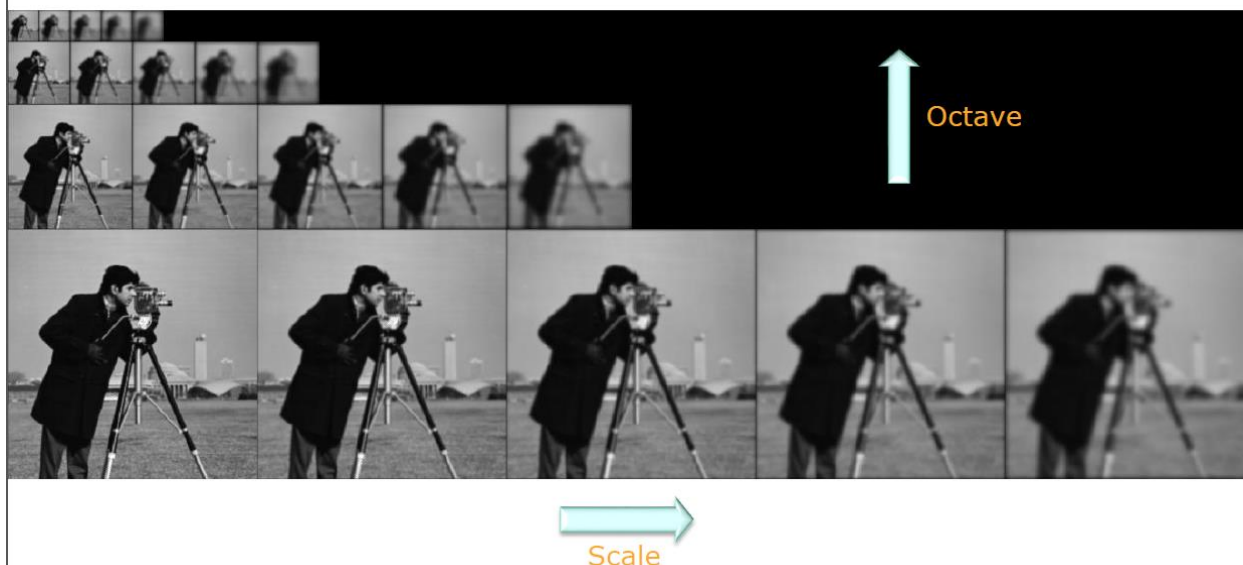
ویژگی های مورد استفاده

SURF ویژگی

در این روش برای استخراج ویژگی سعی بر استخراج نقاط کلیدی تصویر و توصیف آن ها می کند. نقاط کلیدی را با الهام از SIFT و براساس نقاط خاص که در octav ها و scale های مختلف تصویر که octav ها نشان دهنده تصویرهای down

sample شده و همچنین scale ها نشان دهنده تصاویر کانووال شده با گوسی های با پارامترهایی با ضرایب مختلف از یک پایه توصیف می کند .

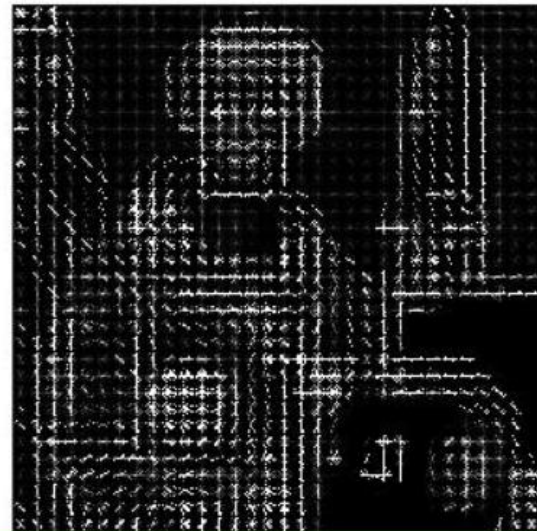
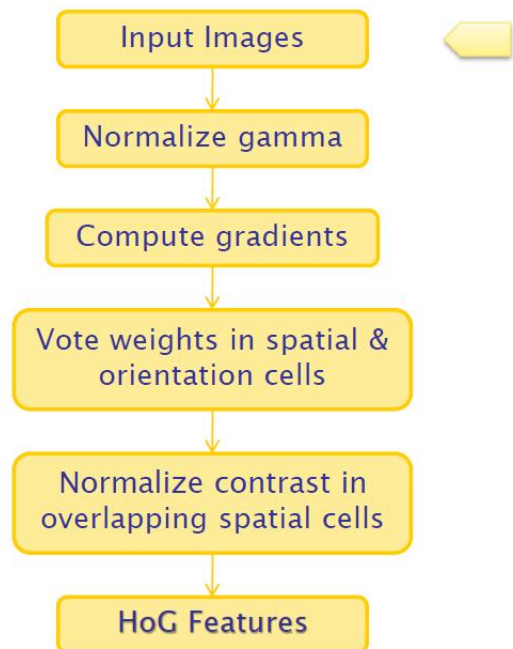
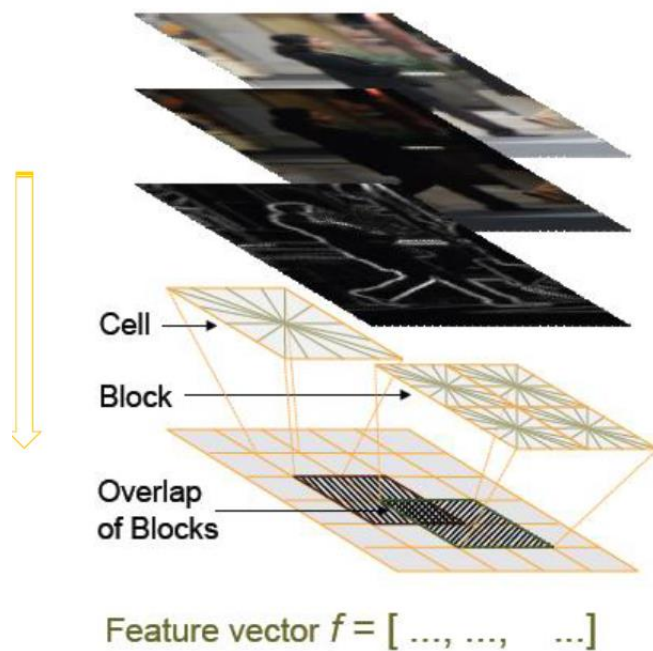
- Gaussian-smoothed images in different octaves



از scale ها مختلف گوسی های تفاضلی (DoG) را محاسبه می کند و در DoG ها اگر نقطه ای نسبت به DoG های اطراف و همچنین Dog متناظر در octav بالایی نقطه خاصی بود آن نقطه را به عنوان نقطه کاندید در می آورد . همچنین برای توصیف نقطه از جمع تقریبی ویژگی های محلی هار در اطراف محل در SURF استفاده می شود. هر توصیف شامل یک آرایه 64 یا 128 تا float که قابل انتخاب است بیان می شود. می توان توصیف نقاط را به صورت تغییر ناپذیر با چرخش و تغییر پذیر با چرخش تعریف کرد. در این پروژه ما از نسخه 64 تایی برای توصیف و تغییر پذیر با چرخش استفاده کردیم؛ علاوه بر این زاویه و اندازه ناحیه توصیف نیز به عنوان داده آموزشی به داده 64 تایی توصیف ناحیه افزوده شد. اما در مورد نحوه استفاده از این ویژگی این طور باید گفته که در مسئله مورد نظر، برای هر عکس تعداد متفاوتی ویژگی SURF بدست می آید که روش استفاده شده به این صورت است که از کنار هم قرار دادن ویژگی های SURF تمام مجموعه عکس های اصلی، مجموعه داده جدیدی می سازیم؛ در واقع اگر عکس اول شامل 40 ویژگی باشد، 40 داده با برجسب عکس اول تولید خواهیم نمود و به همین صورت ادامه خواهیم داد. بنابراین واحد داده آموزشی، یک ویژگی SURF خواهد بود. حال جهت آزمایش هر عکس، به ازای هر ویژگی عکس مورد نظر، توسط مدل مورد نظر برجسب را پیش بینی نموده و بین برجسب ها رای حداکثری گرفته و برجسب نهایی اعلام خواهد شد.

ویژگی HoG

HoG یک ویژگی است که بیشتر برای تشخیص شی در تصویر استفاده می شود. این ویژگی به خوبی می تواند شکل و ظاهر محلی را توصیف کند



دسته بندها و خوشه بندها

در ادامه دسته بندهای مورد استفاده را شرح خواهیم داد که به منظور آزمایش تنها از ۷۵۰۰ داده استفاده شده است و دقت ها بر اساس این تعداد میباشد.

دسته بند KNN

این دسته بند بر مبنای داده های آموزش فضایی را می سازد و بر حسب نزدیکی معیار فاصله اقلیدسی n داده نزدیک را انتخاب و با استفاده از رای گیری بر حسب آن، خروجی مورد نظر را پیش بینی می کند. علت استفاده از این مدل را این طور میتوان بیان نمود که به دلیل ماهیت داده های آموزشی SURF که دارای خاصیت محلی بودن هستند، در نگاه اول مدل KNN که بر اساس داده های محلی داده آزمون عمل میکند، میتواند گزینه مناسبی باشد. در یک قسمت این پروژه از KNN استفاده کردیم و پارامتر n را ۵۰ قرار دادیم که دقت ۰,۸۸۰۸ به دست آمد.

دسته بند SVM

استفاده از این مدل بدون کرنل گاوسی به همراه ویژگی SURF بر خلاف انتظار تنها دقت ۰,۴۳ و با استفاده از کرنل گاوسی دقت ۰,۱۲ را نتیجه داد که بسیار کم بود.

دسته بند RandomForest

این دسته بند از حوزه دسته بند ensemble bagging است. در این دسته بند چند درخت تصمیم و هر درخت با زیر مجموعه ای از مجموعه داده اصلی که با استفاده از نمونه برداری (با جایگزینی و به اندازه مجموعه اصلی) آموزش دیده و سپس داده جدید به همه دسته بند ها داده شده و با اکثریت گرفتن، برچسب داده مشخص می شود. در واقع نقطه قوت این مدل جلوگیری از overfit کردن مدل است. این کار نیز با استفاده از نمونه برداری از مجموعه داده اصلی صورت میگیرد چراکه هر قدر هم یک درخت بر روی مجموعه داده نمونه برداری شده overfit کند، روی داده اصلی نمیتواند overfit کند.

دسته بند اصلی ما در این پروژه Random Forest است. با آموزش این مدل با استفاده از ویژگی SURF بدون هیچ تنظیم پارامتری (تعداد ۱۰ درخت پیش فرض) دقت ۰,۹۵ را بدست آوردیم که با توجه به این نتیجه به عنوان دسته بند پایه انتخاب شد.

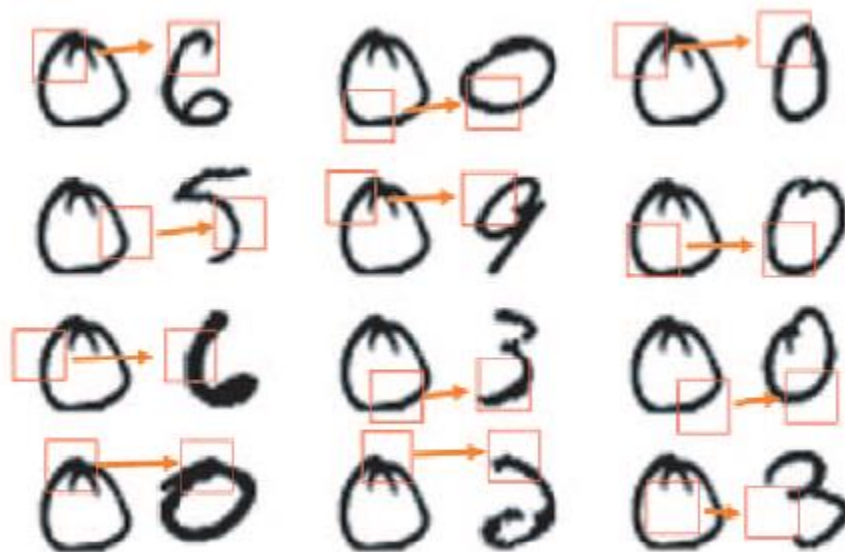
خوشه بند GMM

در این خوشه بند توزیع داده ها را بر مبنای جمع چند توزیع گاوسی می پندارد و سعی می کند با روش تکرار مکرر میانگین داده و واریانس این توزیع ها را بیابد. این مدل در مقایسه با خوشه بند KMeans به دلیل داشتن ساختن خوشه های با شعاع متفاوت دارای قدرت بیشتری بوده که بنابراین این مدل انتخاب شد.

یادگیری با نظارت بر اساس ویژگی SURF با استفاده از پنجره گذاری و ضرورت آن:

هر کدام از قسمت های تصویر در این مجموع داده ویژگی های خاصی دارند که مربوط به ویژگی های آن ناحیه است و اگر با خارج از آن ناحیه مقایسه شود به شدت بر روی کارایی دسته بند تاثیر می گذارد. در واقع اگر داده ها در پنجره ها تقسیم

نشوند، امکان دارد یک ویژگی مربوط به بالای عکس با یک ویژگی مربوط به پایین عکس مقایسه شود که ذاتا مقایسه صحیحی نمیباشد. شکل پایین به خوبی این مشکل را به نمایش می گذارد.

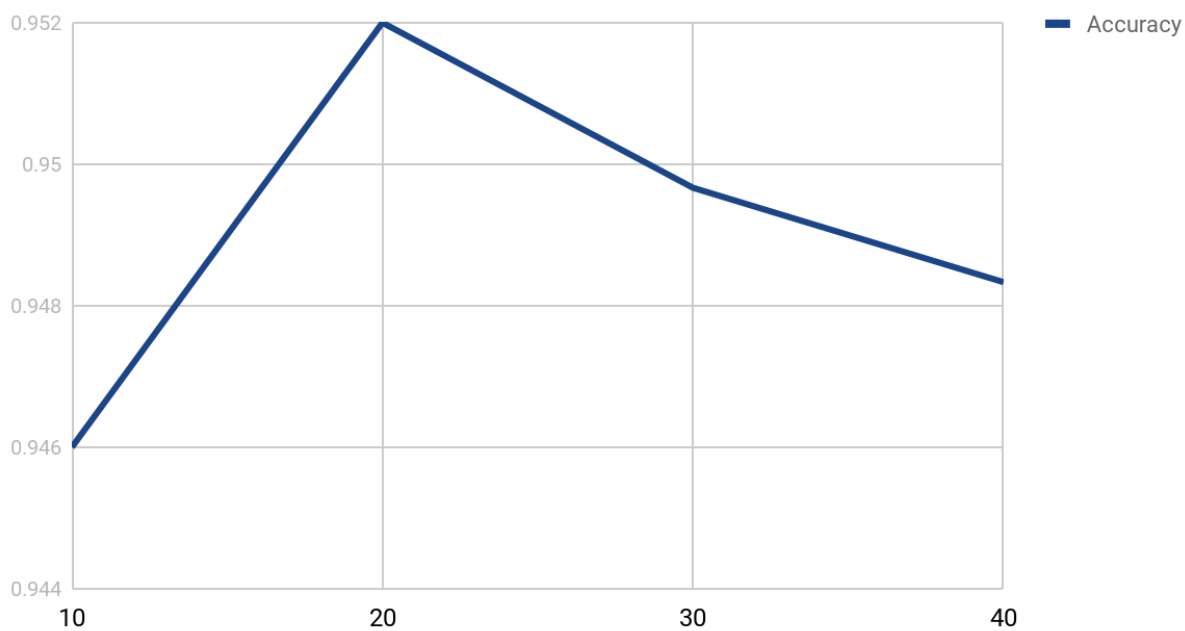


برای فائق آمدن بر این مساله ابتدا ویژگی های تمام تصاویر را بدست آورده (اگر بخواهیم ویژگی را پس از اعمال پنجره بدست آوریم، امکان دارد ویژگی هایی از دست برود) و تعداد مشخصی پنجره با همپوشانی در نظر میگیریم؛ برای هر پنجره یک مدل آموزش خواهیم داد که داده های آموزشی هر پنجره از ویژگی هایی بدست خواهد آمد که مکان آن ها مربوط به همان پنجره است. جهت آزمایش نیز برچسب هر ویژگی عکس با استفاده از مدل هایی پیش بینی خواهد شد که مکان آن با مکان پنجره مدل منطبق باشد. در واقع میتوان این طور ادعا کرد که مدل ارائه شده نوعی مدل ensemble میباشد. در مدل توضیح داده شده از سه مدل KNN، SVM و RandomForest جهت آموزش هر پنجره استفاده شد که بهترین نتیجه با استفاده از RandomForest بدست آمد.

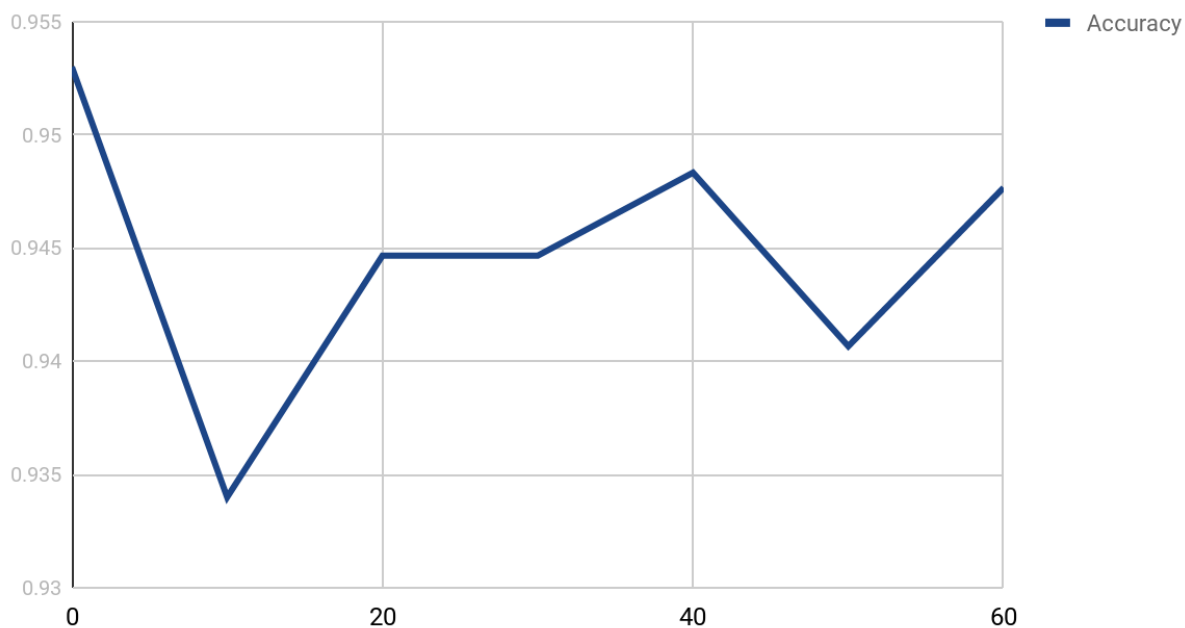
توضیح روش دسته بندی پیاده شده :

ابتدا تصاویر را با افزودن ۱۰ سطر و انتها و همچنین ده ستون به ابتدا و انتها به تصویر $48 * 48$ رساندیم و سپس تصویر را در 4 scale کردیم و سایز هر تصویر $192 * 192$ شد. لزوم این کار برای استخراج ویژگی SURF بود. سپس ویژگی های SURF 64 تایی و متغیر با چرخش هر تصویر و مکان هایشان را استخراج کردیم و سپس این داده ها به پنجره های مختلف که در این مساله $8 * 8$ پنجره بدون هم پوشانی بود تقسیم کردیم و دسته RandomForest هر بخش را آموزش دادیم. در فاز در آوردن دسته داده ورودی بعد از استخراج ویژگی های SURF تصویر ویژگی های هر پنجره را به دسته بند آن پنجره می فرستیم و دسته هر ویژگی را بدست می آوریم و سپس اکثریت می گیریم و دسته آن داده را بدست می آوریم. به منظور تعیین پارامتر ها با استفاده از روش KFoldCrossValidatoin، مدل را روی تنها 300 داده آموزش دادیم و با k برابر ۵ آزمایش نمودیم.

Number of Estimators



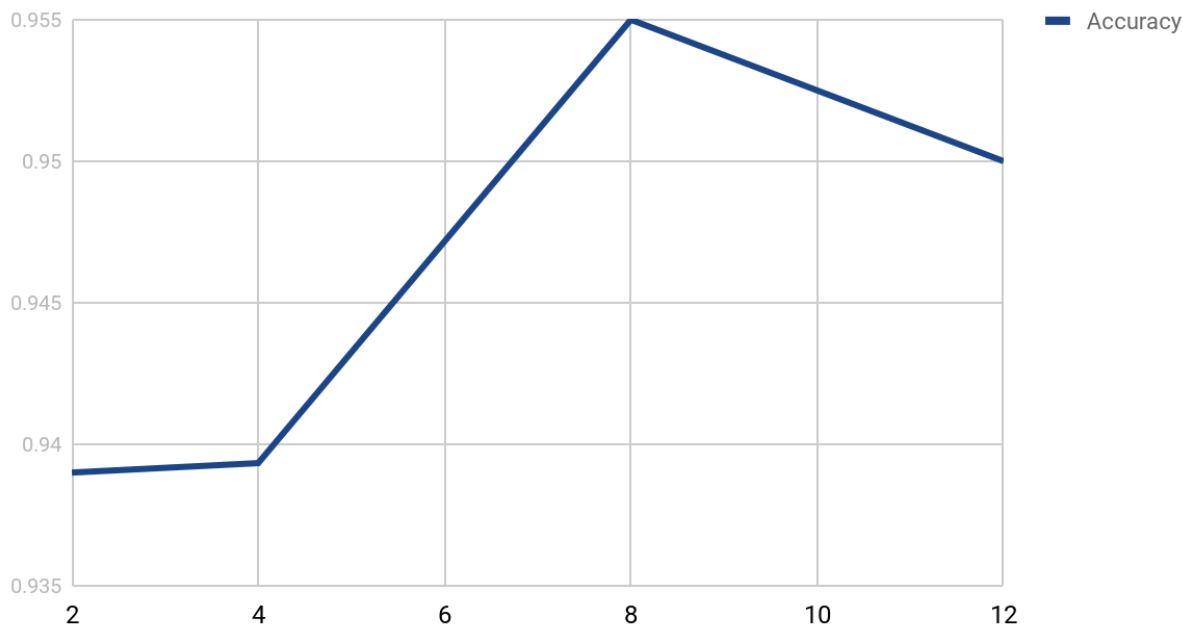
Number of PCA Components



منظور از صفر بدون اعمال کاهش بعد است.

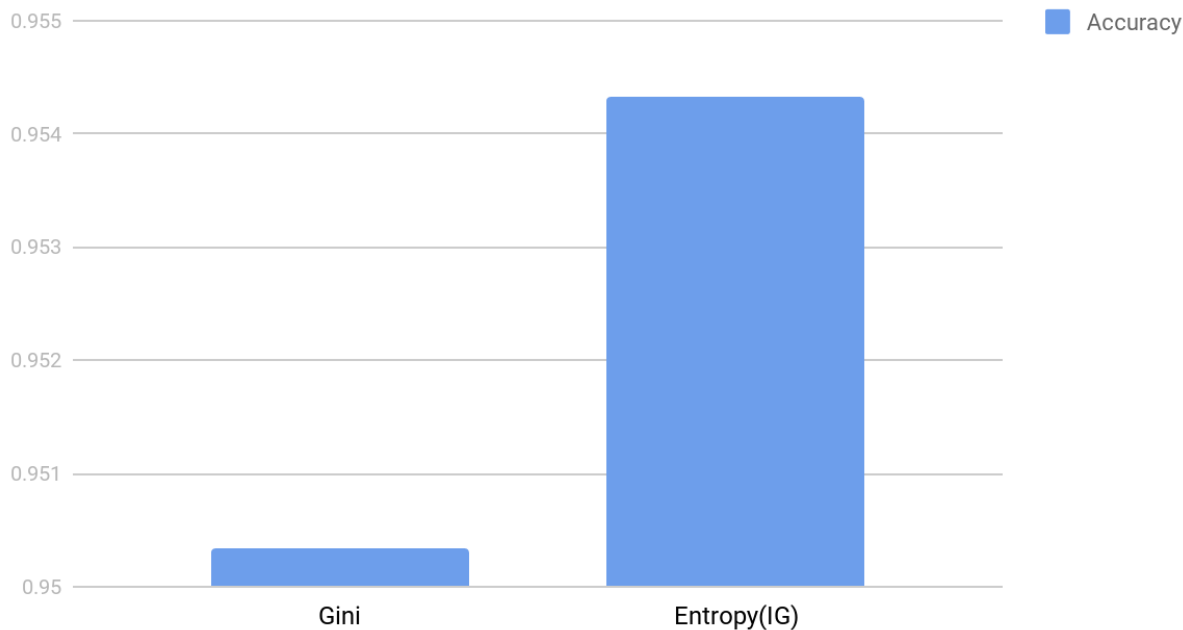
لازم به ذکر است که در این روش به دلیل استفاده از ویژگی SURF که در نهایت تنها ۶۴ ویژگی خواهیم داشت که با استفاده از اعمال PCA (که بهتر از LDA عمل نمود) دقت به ۹۴ درصد کاهش یافت؛ بنابراین از کاهش بعد استفاده نشد.

Number of Windows



برای تعداد پنجره تعداد پنجره زوج انتخاب شد چرا که در غیر این صورت یک پنجره دقیقاً در وسط عکس قرار میگیرد و بیشتر ویژگی ها در آن قرار خواهد گرفت که برای جلوگیری از این بایاس، تعداد پنجره ها زوج در نظر گرفته شد.

Decision Tree Cirterion



البته با این وجود که این معیار عملکرد بهتری داشته به دلیل سریع تر بودن معیار Gini از آن استفاده شد.
بهترین تعداد پنجره : $8 * 8$

بهترین تعداد ابعاد کاهش یافته در کاهش بعد(بدون در نظر گرفتن کاهش بعد ندادن): ۴۰

بهترین تعداد درخت: ۲۰

بهترین معیار در درخت تصمیم: Entropy(Information Gain)

با این روش توانستیم به دقت ۰,۹۸۱۱ که به تفکیک هر دسته

number 0 accuracy: 1.0

number 1 accuracy: 0.9955947136563876

number 2 accuracy: 0.9922480620155039

number 3 accuracy: 0.9900990099009901

number 4 accuracy: 0.9898167006109979

number 5 accuracy: 0.9607623318385651

number 6 accuracy: 0.9822546972860126

number 7 accuracy: 0.9659533073929961

number 8 accuracy: 0.9794661190965093

number 9 accuracy: 0.9514370664023786

دست یافتیم.

در مقاله ای که به عنوان پایه کار قرار داشت از knn استفاده کرده بود و دقت را حدود ۰,۹۶ بدست آورده بود که ما با تغییر بعضی از پارامترها و استفاده از Random Forest توانستیم دقت را به بیش از ۰,۹۸ برسانیم.

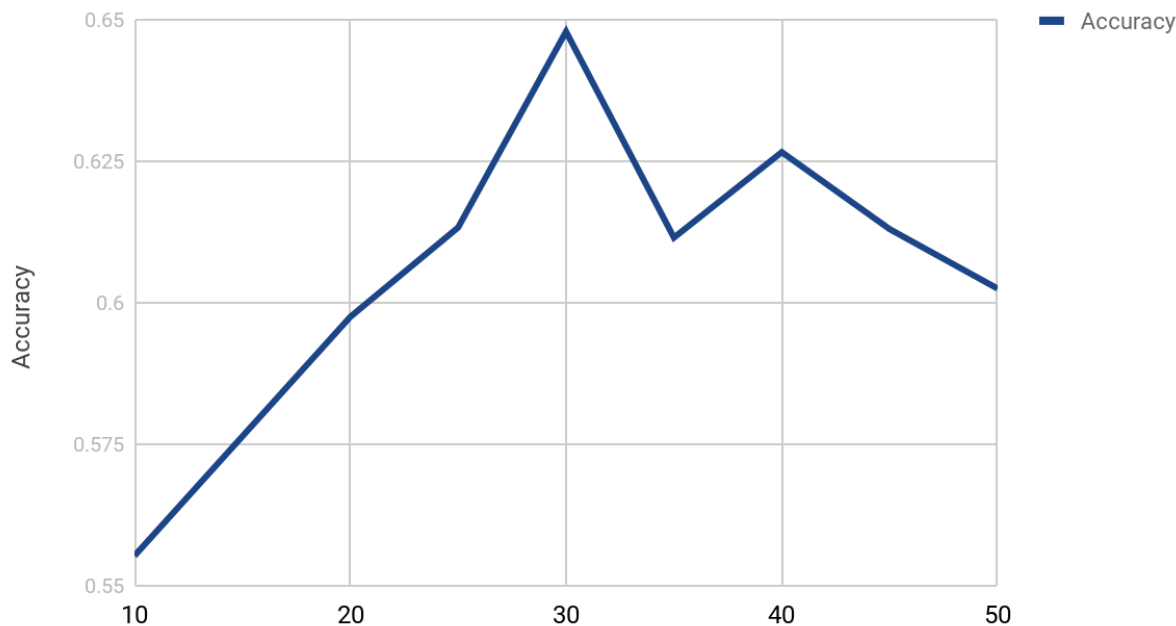
توضیح روش خوشه بند پیاده شده :

در این ویژگی های HoG با پارامترهای

```
winSize = (28,28)
blockSize = (8,8)
blockStride = (4,4)
cellSize = (2,2)
nbins =7
derivAperture = 1
winSigma = 4.
histogramNormType = 0
L2HysThreshold = 2.0000000000000001e-01
gammaCorrection = 0
nlevels = 7
```

که با سعی و خطا متوجه شدیم بهترین پارامترها هستم استخراج کردیم. سپس ویژگی ها را به pca دادیم و با تغییر تعداد کامپوننت pca متوجه شدیم که بهترین تعداد ۲۹ می باشد که با این ترکیب توانستیم به $\text{purity} = ۰,۶۵۶۵$ و $\text{adjust} = ۰,۴۶۵۱۲۷۰۲۶۰۲۰۱۸۶۴$ دست یابیم.

Accuracy



خوشه بندی که نتیجه نداد!

از آنجایی که با استفاده از پنجره گذاری در بخش دسته بندی به نتیجه قابل قبولی دست یافتیم، قصد داشتیم تا با استفاده از این تکنیک در خوشه بندی نیز به نتیجه مناسبی دست یابیم. به این منظور برای هر پنجره یک خوشه بند آموزش داده میشود؛ اما مشکل آنجاست که برجسبی وجود ندارد تا خوشه های متناظر پنجره های مختلف را بیابیم؛ به عبارت دیگر باید روشی جهت یافتن نگاشتی بین خوشه های پنجره ها میافتیم. به این منظور روش های متفاوتی وجود دارد که یک مورد از آن ها را در ذکر خواهیم کرد.

به این منظور میتوان گرافی از خوشه ها ساخت و با پردازش بر روی آن به نگاشت مورد نظر دست یافت. در این گراف هر گره یک خوشه یک پنجره و یالها میزان شباهت بین خوشه ها است. در نهایت باید این گراف را به تعداد خوشه های اصلی زیر گراف بدون اشتراک هم اندازه با کمترین تغییر در یالهای گراف تبدیل نمود. حال هر زیرگراف نماینده یک خوشه در خوشه بند اصلی خواهد بود. به مسئله فوق k -way uniform graph partitioning گفته میشود که مسئله ای NP-Complete است که برای حل آن از هیوریستیک های متفاوتی استفاده میشود.

در مورد مسئله مورد نظر، ما نیز از یک هیوریستیک ساده استفاده نمودیم. میدانیم بیشتر ویژگی ها در میانه تصویر وجود دارند؛ بنابراین احتمالا خوشه های این پنجره میانی، صحیح ترین خوشه بندی را ارائه میکنند. از این پنجره شروع کرده و به ازای هر پنجره ی دارای محدوده مشترک با آن، برجسب خوشه به ازای داده های مشترک دو پنجره بدست آمده و پر تکرار ترین برجسب پنجره اول به پر تکرارترین برجسب پنجره دوم نگاشت داده شده و به همین ترتیب ادامه میدهیم و به سراغ سایر پنجره ها خواهیم رفت. با این روش عملا گراف تولید شده به تعداد خوشه های اصلی زیر گراف تقسیم میشود و نگاشت نهایی بدست می آید.

جهت آموزش مدل فوق از ویژگی SURF استفاده شد اما پس از اجرا، بر خلاف آنچه که انتظار داشتیم به purity حداکثر برابر ۰,۲۷۶، با استفاده از PCA با ۱۳ المان دست یافتیم. علت این نتیجه را این طور میتوان توجیه نمود که خوشه بند پنجره میانی که بیشترین تعداد ویژگی ها در آن وجود دارد قادر به خوشه بندی مناسب داده ها نداشت و معیار purity پنجره میانی عددی در حدود ۰,۲۵ بود.

منابع

Uchida, Seiichi, and Marcus Liwicki. "Part-based recognition of handwritten characters." *Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on*. IEEE, 2010.