



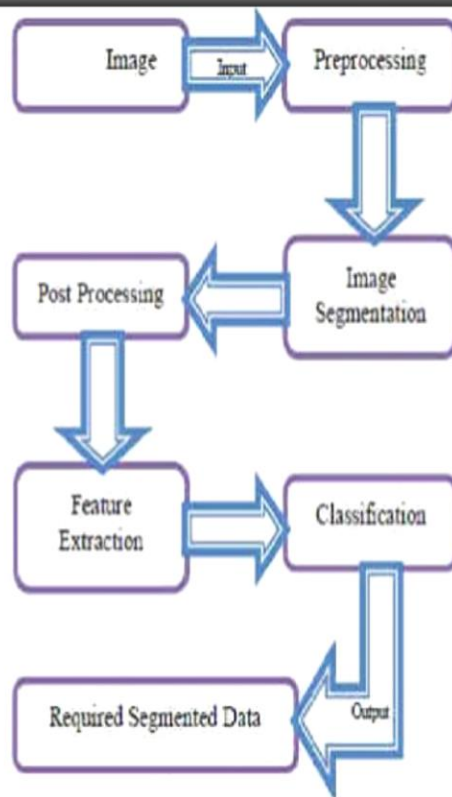
دانشگاه شهید باهنر کرمان

سمینار درس بینایی ماشین

موضوع : پردازش و تشخیص چهره با ماشین لرنینگ با استفاده از تکنیک یادگیری عمیق

پردازش تصویر چیست ؟

- پردازش تصویر یا Image Processing، امروزه به پردازش تصویر دیجیتال گفته می شود که نیازمند دانش رایانه ای است و سیگنال دیجیتالی که توسط دوربین و یا اسکنر برداشته شده را پردازش می کند.
- در پردازش تصویر یک تصویر به عنوان ورودی دریافت میشود و با انجام یک سری عملیات بر روی آن یک تصویر یا یک مجموعه از نشان‌های ویژه یا متغیرهای مربوط به تصویر به عنوان خروجی از آن دریافت می‌شود.



○ خروجی های پردازش تصویر :

○ 1. نمایش و چاپ تصویر.

○ 2. ویرایش تصویر.

○ ۳. بهبود تصویر.

○ 4. کشف و تشخیص یک ویژگی خاص در تصویر.

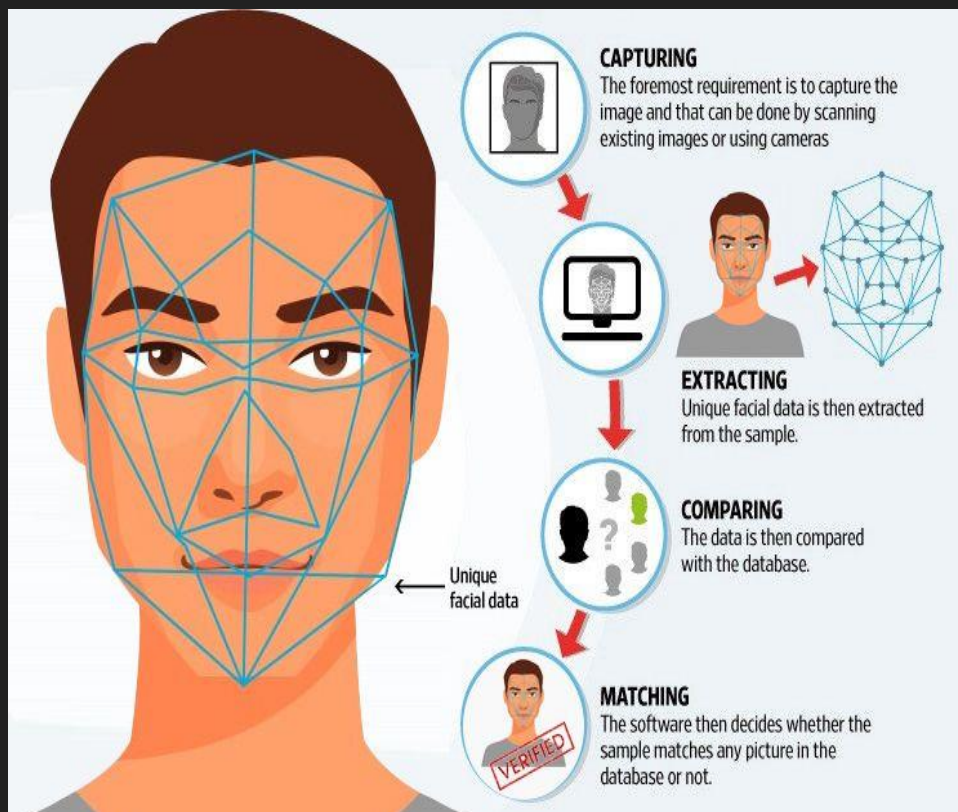
○ ۵. فشرده سازی تصویر.

عملیات های اصلی که در عملیات پردازش تصویر انجام میشوند

- تبدیلات هندسی مثل تغییر اندازه ، چرخش و...
- رنگ مثل تغییر روشنایی، وضوح ویا تغییر فضای رنگ
- ترکیب تصاویر مثل ترکیب دو یا چند تصویر
- فشرده سازی تصویر مثل کاهش حجم تصویر
- ناحیه بندی پرونده یعنی تجزیه تصویر به نواحی با معنی
- بهبود کیفیت پرونده مثل کاهش نویز و افزایش کنتراست
- سنجش کیفیت تصویر
- ذخیره سازی اطلاعات در تصویر
- انطباق تصاویر

شناسایی چهره چیست ؟

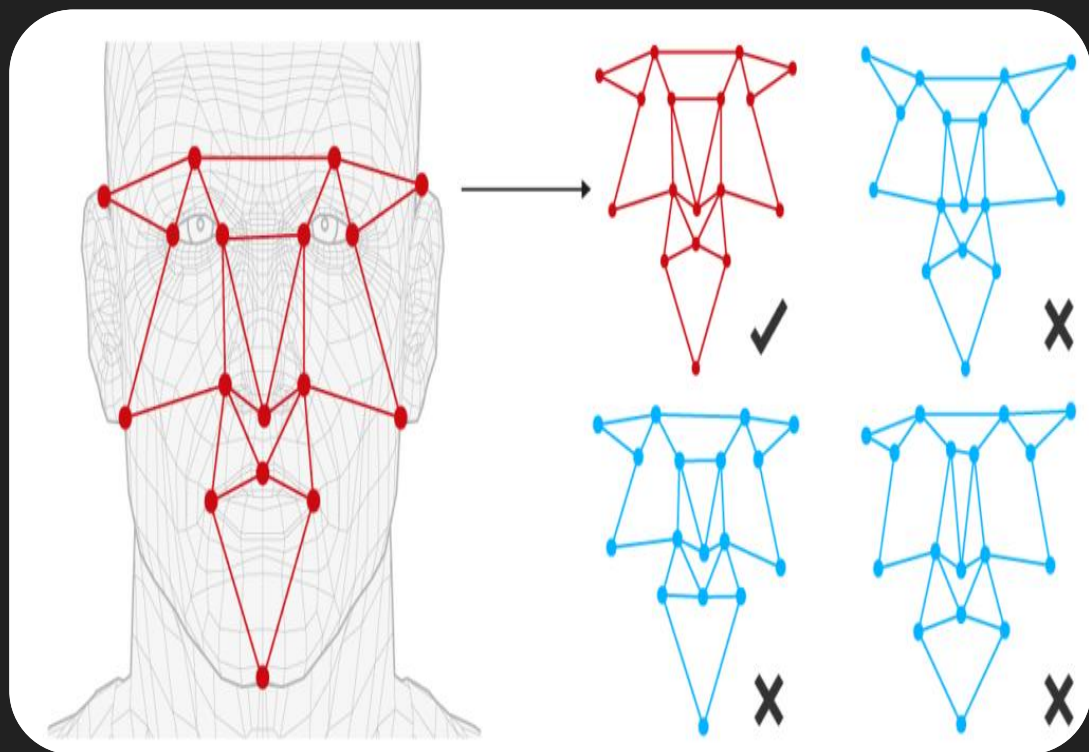
- یک فناوری بیومتریک است که با استفاده از تجزیه و تحلیل الگوهایی که از صورت یک فرد به دست می آید قادر است افراد را از یک دیگر تشخیص دهد و آن ها را شناسایی نماید و بتواند هویت افراد را تایید کند.
- با پیشرفت تکنولوژی از تشخیص چهره در بسیاری از زمینه ها مانند بازگشایی قفل صفحه گوشی تا کاربرد های امنیتی در سازمان های مختلف استفاده میشود .



چگونه کار میکند ؟

- اول از همه باید به سیستم آموزش داده شود که یک چهره چیست تا بتواند آن را از سایر موجودیت های اطرافش تشخیص دهد.
- در قدم بعدی باید به سیستم توانایی تشخیص داده شود. برای انجام این کار روش های متعددی وجود دارد، اما معمولا از یک شبکه عصبی دومی استفاده می شود. در این روش، با استفاده از تصاویر به سیستم آموزش داده می شود که چگونه بتواند یک فرد را از دیگری تشخیص دهد.
- به طور کلی فناوری تشخیص چهره با استفاده از اندازه گیری ده ها ویژگی قابل تشخیص در صورت افراد، می تواند آن ها را شناسایی کند

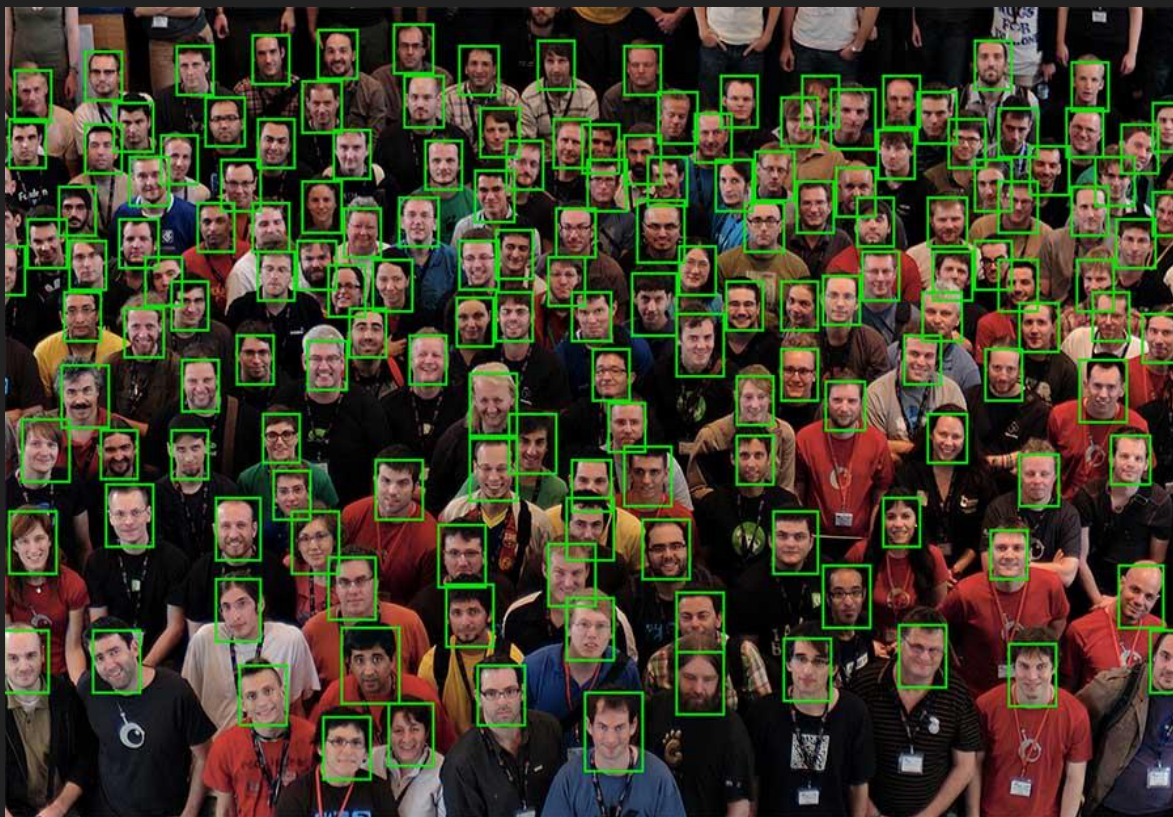
به طور کلی :



1. فناوری تشخیص چهره در ابتدا با استفاده از هندسه یک صورت، یک "اثر صورت" (faceprint) برای هر فرد ایجاد می کند که همانند اثر انگشت منحصر به فرد است.

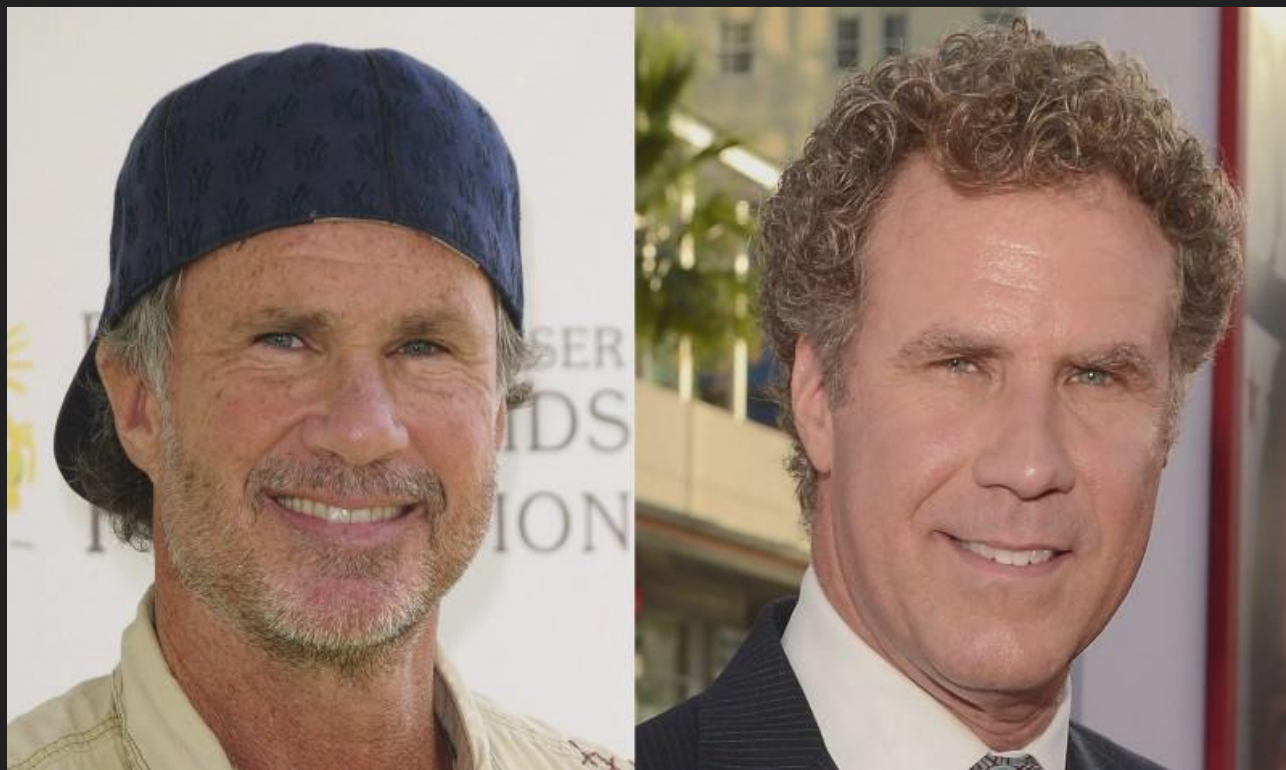
2. از این faceprint برای مقایسه با چهره افرادی که از قبل به سیستم داده شده است استفاده می شود و سیستم بر اساس بیشترین شباهت تصاویر را رتبه بندی می کند. صحت گزینه هایی که رتبه بندی شده اند در نهایت با تایید یک اپراتور انسانی مشخص می شوند.

کاربرد



- امروزه شرکت ها با استفاده از روش های گوناگون سعی در پیشرفته تر کردن این فناوری هستند.
- یکی از این روش ها استفاده از تکنیک یادگیری عمیق میباشد .
- در این روش تعداد زیادی ورودی را به سیستم میدهیم ، سیستم با پارامتر ها و داده هایی که به عنوان ورودی دریافت کرده آموزش میبیند و خروجی مطلوب را به ما میدهد .
- از شرکت های پیشتاز در این عرصه میتوان به فیسبوک اشاره کرد که از این تکنیک برای تگ کردن خود کار افراد استفاده کرده است .

بررسی مرحله به مرحله این روش



○ فرد سمت چپ چاد اسمیث (Chad Smith) موسیقیدان راک و شخص سمت راست ویل فرل (Will Ferrell) بازیگر معروف است. مساله‌ای که در این مطلب قرار است حل شود تشخیص چهره این دو فرد از یکدیگر است.

مسئله تشخیص چهره مجموعه ای از مسائل مرتبط با هم است که مراحل آن به شرح زیر میباشد :

- 1. نگاه کردن به تصویر و شناسایی تمامی چهره های موجود در آن
- 2. تمرکز بر روی هر چهره و توانایی درک اینکه هر چهره حتی اگر در جهتی چرخیده باشد یا نور نا مناسب باشد باز هم متعلق به همان شخص است .
- 3. انتخاب ویژگی هایی که در هر شخص متفاوت است مانند فاصله بین چشم ها ، اندازه بینی و گوش ها ، طول چهره فرد و دیگر موارد .
- 4. مقایسه ویژگی های چهره همه افرادی که در حال حاضر شناخته شده اند با یکدیگر برای تشخیص نام هر فرد.
- تمامی مراحل بالا در کسری از ثانیه در ذهن انسان انجام میشوند .

گام اول : یافتن همه چهره ها



○ اولین گام برای حل مساله مطرح شده، تشخیص چهره است.

○ واضح است که پیش از تشخیص اینکه یک تصویر متعلق به چه کسی است، باید خود چهره‌ها در تصویر تشخیص داده شوند.

○ در اغلب دوربین‌هایی که از ده ساله گذشته تاکنون مورد استفاده قرار گرفته‌اند، توانایی تشخیص چهره به شکلی که در تصویر نشان داده شده وجود دارد.

○ هنگامی که دوربین به‌طور خودکار چهره را تشخیص می‌دهد، می‌تواند اطمینان حاصل کند که همه چهره‌ها در مرکز توجه دوربین (فوکوس) قرار دارند.

○ از این قابلیت می‌توان برای اهداف متعدد دیگری مانند کشف نواحی از تصویر که قرار است به گام بعد انتقال یابند نیز استفاده کرد.

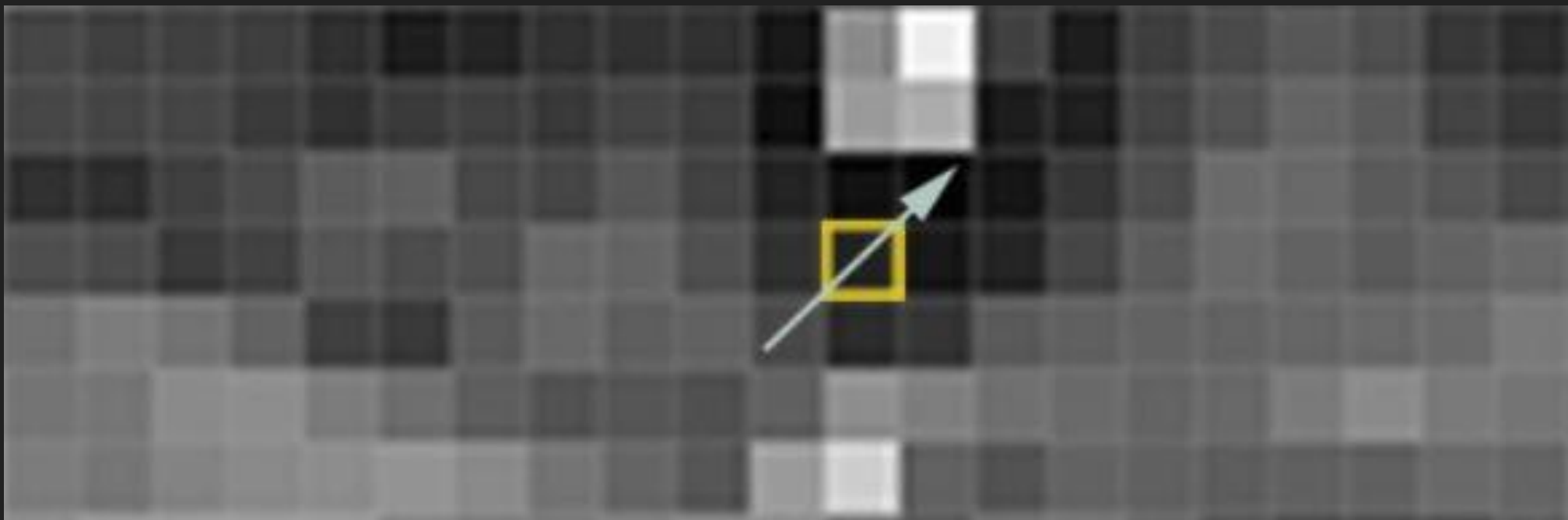


- در این مطلب، از روشی برای تشخیص چهره استفاده شده که در سال ۲۰۰۵ ابداع و به آن «هیستوگرام شیب‌گرا» (Histogram of Oriented Gradients-HOG) گفته می‌شود.
- برای آغاز کار تشخیص چهره در یک تصویر، ابتدا باید آن را سیاه و سفید کرد. زیرا برای این کار نیازی به داده‌های رنگی نیست.

- سپس، کلیه پیکسل‌های موجود در تصویر به یکباره مورد بررسی قرار می‌گیرند. برای هر پیکسل مجزا، باید پیکسل‌هایی که مستقیماً در اطراف آن قرار دارند بررسی شوند.
- هدف، کشف این است که یک پیکسل در مقایسه با پیکسل‌هایی که اطراف آن را احاطه کرده‌اند چقدر تیره‌تر است.

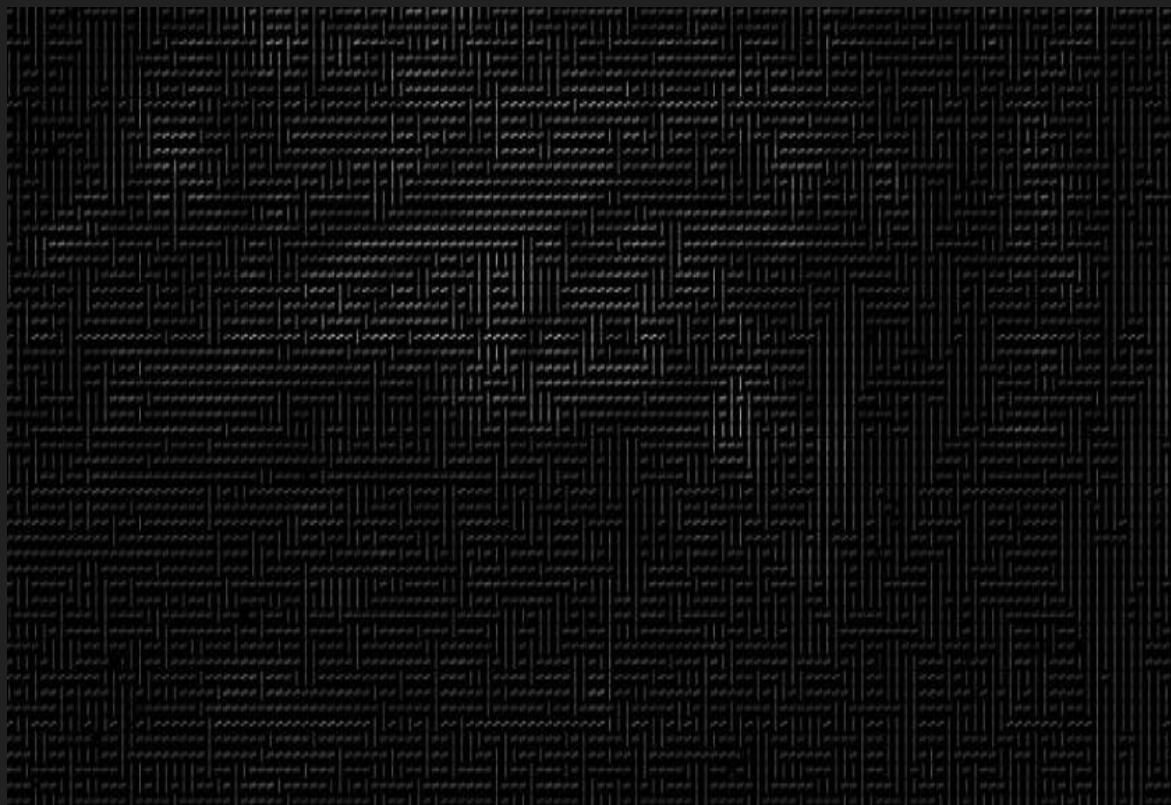


○ قدم بعدی ، ترسیم پیکانی است که نشان می‌دهد پیکسل‌ها در کدام جهت تیره‌تر می‌شوند.



○ اگر این فرآیند برای هر پیکسل مجزا در تصویر انجام شود، می‌توان هر پیکسل را با یک پیکان جایگزین کرد. این پیکان‌ها «گرادیان» (gradient) نامیده می‌شوند و در کل تصویر از سمت روشن به تیره‌تر هستند. بنابراین، گرادیان‌ها جهت روشنی به تیرگی را در تصویر نمایش می‌دهند.

○ ذخیره‌سازی گرادیان‌ها برای هر پیکسل مجزا، جزئیات زیادی را برای تحلیل به ارمغان می‌آورد. در صورتی که بتوان جهت جریان روشنایی/تاریکی را در سطح بالاتر دید، می‌توان الگوی پایه‌ای تصویر را تشخیص داد.



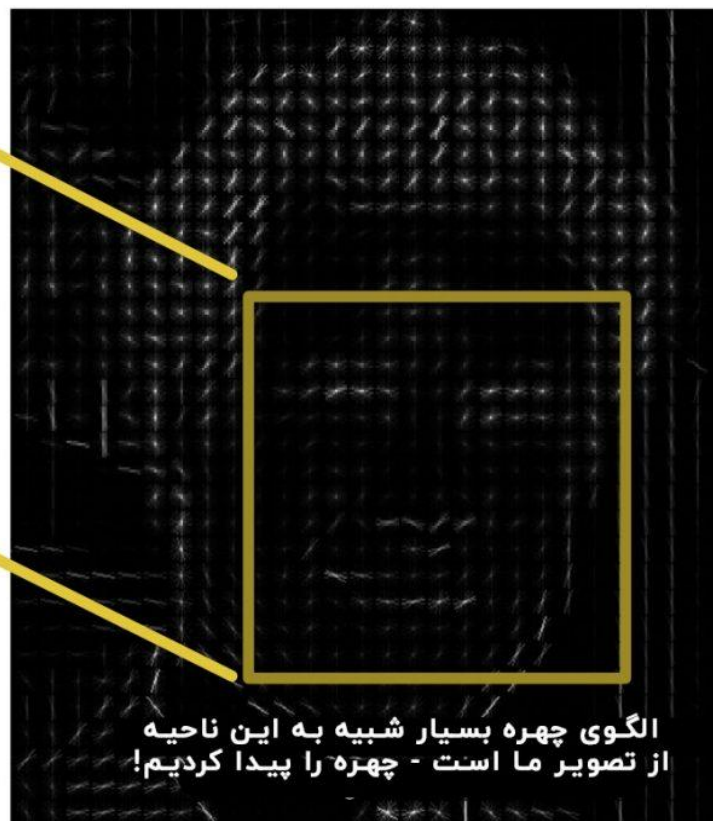


○ عکس اصلی به HOG تبدیل شد که ویژگی های مهم و اصلی از عکس را علی رغم روشنایی عکس دریافت می کند.



نسخه ی HOG تصویر ما

HOG الگوی چهره ها که
با استفاده از تصاویر زیادی
از چهره ها تولید شده است



الگوی چهره بسیار شبیه به این ناحیه
از تصویر ما است - چهره را پیدا کردیم!

- برای پیدا کردن صورت ها در این عکس HOG ، کاری که باید انجام دهیم این است که یک قسمت از تصویرمان که بیشترین شباهت به الگوی شناخته شده HOG که از میان دیگر صورت هایی که برای آموزش استفاده شده اند، استخراج شده است را دارد، پیدا کنیم



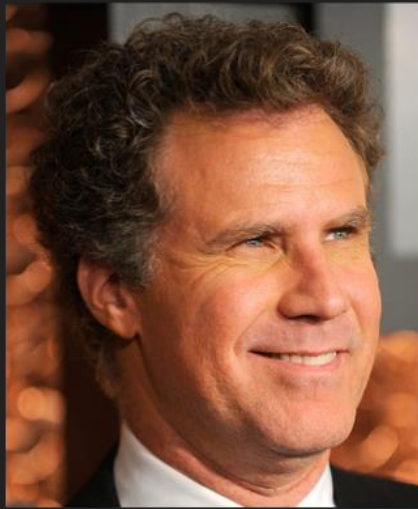
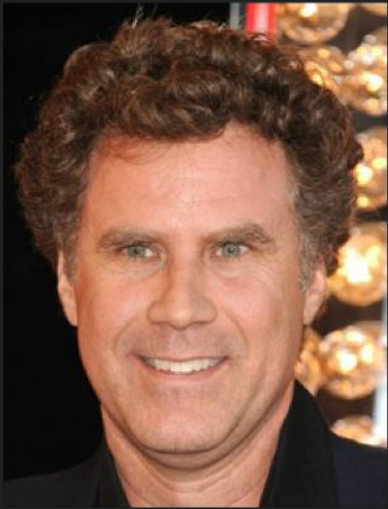
○ حالا میتوانیم چهره ها را در هر تصویر پیدا کنیم

شناسایی چهره



- سخت ترین مرحله تفکیک چهره ها از یکدیگر است .
- ساده ترین رویکرد : مقایسه همه عکس ها با هم .
- نیاز داریم راهی برای استخراج تعدادی معیار اندازه گیری در هر چهره، پیدا کنیم.
- سپس می توانیم چهره ناشناس را با همان روش معیار سنجی کنیم و نزدیکترین چهره شناخته شده را با معیار ها پیدا کنیم.

برای چهره های زاویه دار و متمایل باید چه کاری انجام داد؟

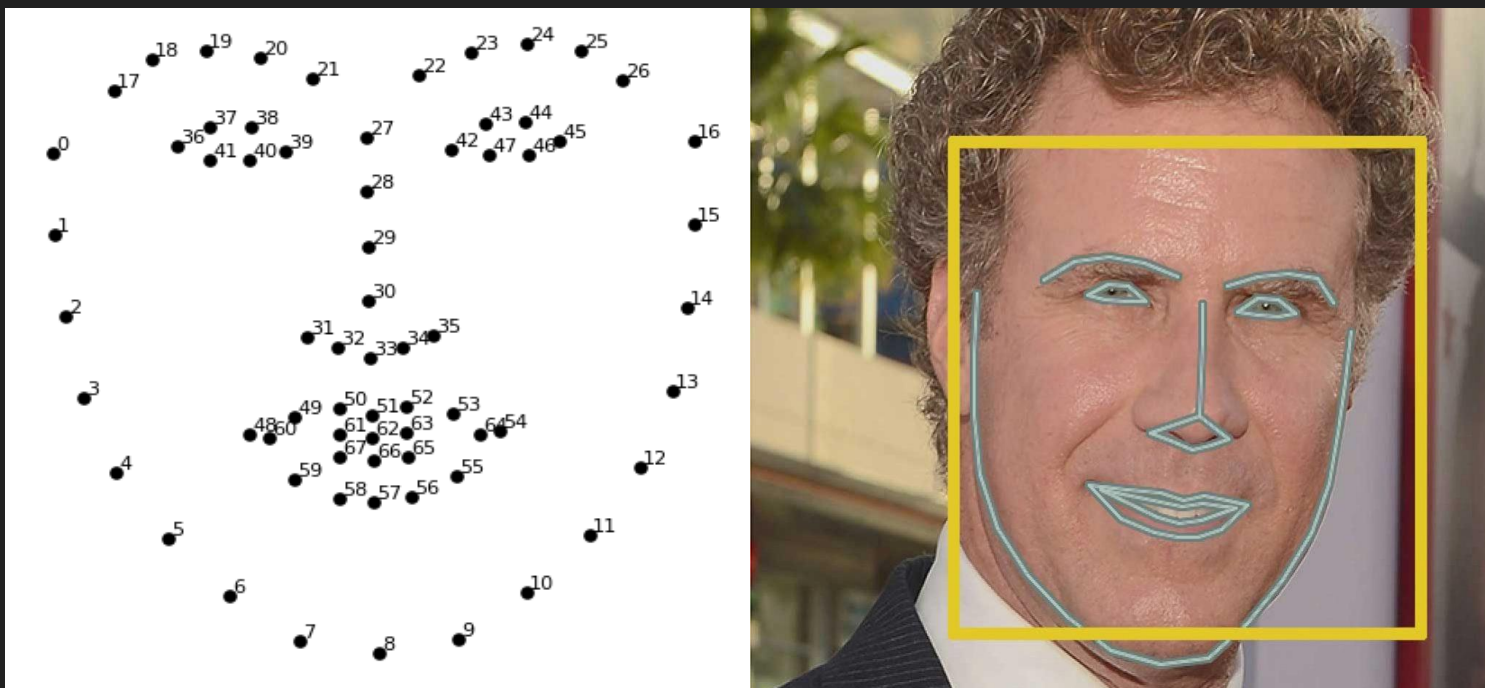


○ چشم انسان به راحتی میتواند تشخیص دهد که هر دو عکس متعلق به Will Ferrell است . ولی کامپیوتر ها ممکن است این تصاویر را کاملا متفاوت از یکدیگر تشخیص دهند. حال با تکنیک هایی تغییراتی را در عکس ایجاد میکنیم که اجزای اصلی مانند چشم بینی و دهن همیشه در محلی مشخص قرار گیرند

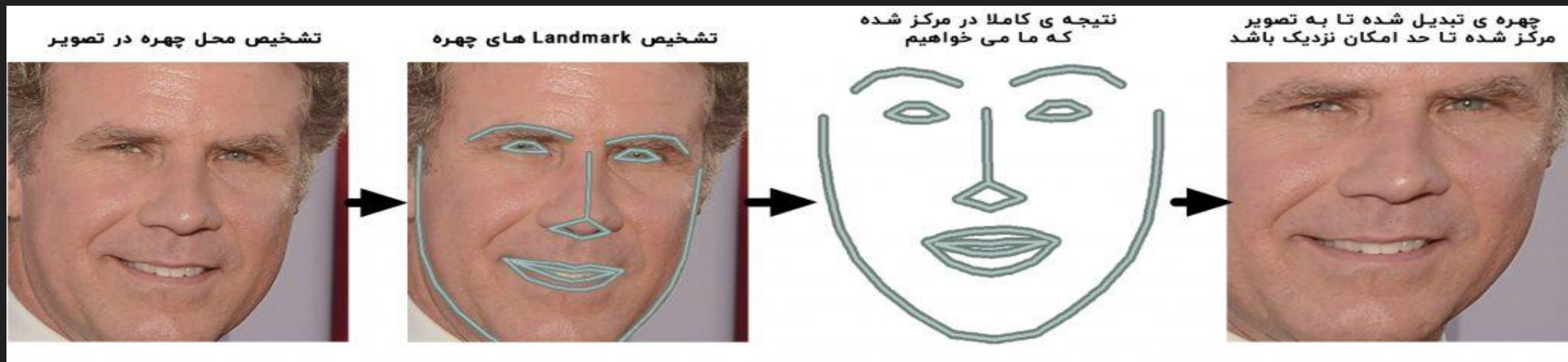
○ برای انجام این کار، ما قرار است از یک

الگوریتم به نام Face Landmark Estimation استفاده کنیم.

○ نظریه اساسی آن این است که ما ۶۸ نقطه‌ی خاص در عکس می‌یابیم (که Landmark نام دارند) و در هر چهره‌ای وجود دارند - بالای چانه، حاشیه‌ی بیرونی هر چشم، حاشیه‌ی درونی هر ابرو و غیره. سپس ما یک الگوریتم یادگیری ماشین را آموزش خواهیم داد تا قادر باشد این ۶۸ نقطه‌ی اختصاصی را در هر چهره‌ای پیدا کند.



- حالا که می دانیم چشم ها و دهان کجا قرار گرفته اند، ما به راحتی تصویر را چرخش می دهیم، برایش مقیاس تعیین می کنیم و برش می دهیم، در نتیجه چشم ها و دهان در بهترین محور و حالت ممکن قرار می گیرند. ما هیچ گونه پیچ و تاب ۳ بعدی روی عکس انجام نخواهیم داد زیرا آن ممکن است در عکس اعوجاج ایجاد کند. ما فقط قرار است از دگرگونی تصویری اصلی مانند چرخش یا مقیاس که از خط های موازی جلوگیری می کنند، استفاده کنیم

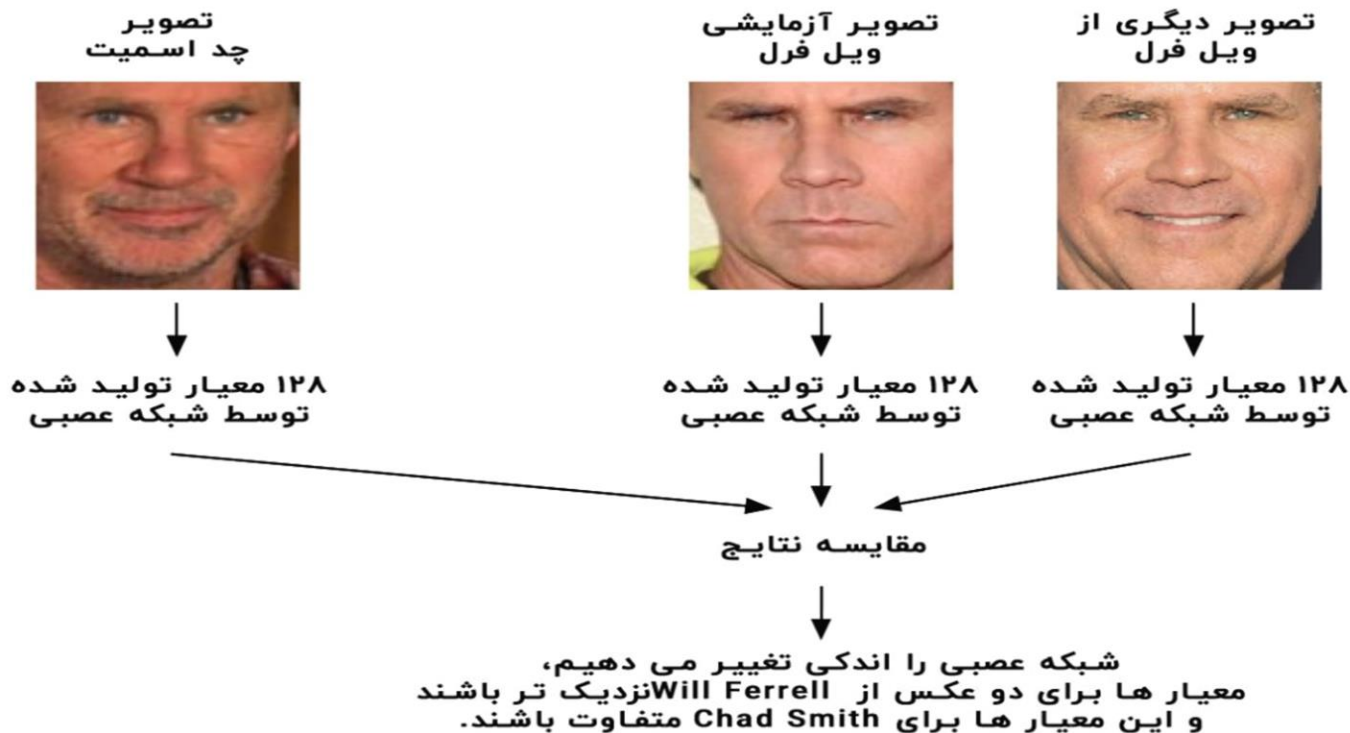


قابل اطمینان ترین راه برای معیار سنجی یک صورت

- معیار هایی که برای ما انسان ها واضح می باشند، (مانند رنگ چشم) برای یک کامپیوتر که به هر پیکسل را جداگانه بررسی می کند، منطقی به نظر نمی رسد.
- محققان کشف کرده اند که دقیق ترین رویکرد این است که به کامپیوتر اجازه دهیم تا معیار هایی که نیاز دارند جمع آوری کنند را خودشان کشف کنند.
- یادگیری عمیق در فهمیدن اینکه چه قسمت هایی از چهره برای معیار سنجی مهم تر اند ، بهتر از انسان عمل می کند.
- راه حل این است که یک شبکه عصبی عمیق کانولوشنی آموزش دهیم که قرار است آن را آموزش دهیم تا ۱۲۸ معیار برای هر چهره تولید کند.

روند آموزشی به این صورت است :

یک مرحله آموزش "سه گانه":



1. بارگیری یک تصویر صورت از فرد شناخته

شده

2. بارگیری یک تصویر دیگر از همان فرد

شناخته شده

3. بارگیری یک تصویر از یک فرد کاملاً

متفاوت

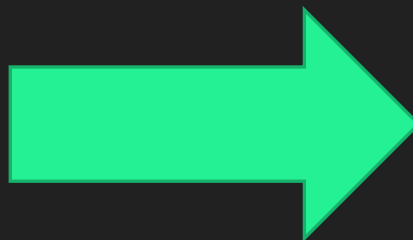
- بعد از میلیون ها بار تکرار این مرحله برای میلیون ها تصویر از هزاران فرد مختلف، شبکه عصبی یاد می گیرد تا به طور امن و قابل اطمینان، ۱۲۸ معیار برای هر شخص تولید کند. ۱۰ تصویر مختلف از یک شخص یکسان، تقریباً باید معیار های یکسانی بدهد.
- به ۱۲۸ معیار برای هر چهره Embedding می گویند.



شبکه عصبی ترین شده برای تولید Embedding

- این روند از آموزش یک شبکه عصبی کانولوشنی برای اینکه Embedding صورت را استخراج کند، به مقدار زیادی داده و توان کامپیوتری نیاز دارد
- حتی با یک کارت گرافیک NVidia Tesla گران قیمت نیز نیاز به ۲۴ ساعت آموزش مکرر دارد تا به دقت خوب برسد.
- ولی همین که یکبار شبکه آموزش ببیند، حتی برای چهره هایی که قبلا ندیده است می تواند معیار هایی تولید کند. این مرحله فقط به یک بار انجام نیاز دارد.

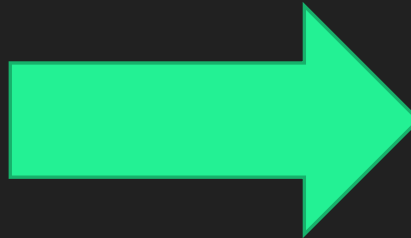
نمونه Embedding



Gholamreza Encoding

```
[ -0.13571851 -0.01698408  0.03395299  0.00245783 -0.02608508 -0.07483841  
  0.1066476  -0.02763853  0.23058626 -0.04604723  0.17017962 -0.00655842  
 -0.17140265 -0.04825669 -0.07843701  0.06690523 -0.07184006 -0.10327367  
 -0.06793824 -0.09690443  0.10108522  0.01482281 -0.02181487  0.07144096  
 -0.11482926 -0.29705817 -0.08958135 -0.18997549  0.09628405 -0.1154581  
  0.05107953  0.05847457 -0.13861778 -0.10231259  0.00561826  0.05975875  
  0.04353987 -0.04710814  0.25628752  0.02616877 -0.11835837 -0.00987772  
  0.01333191  0.31040645  0.0307086  0.09023419  0.0072752  -0.05080656  
  0.09526203 -0.22579312  0.08662529  0.09826126  0.08112846  0.02328907  
  0.08069704 -0.18838578  0.06840493  0.03349276 -0.27691573  0.14384435  
  0.04585487 -0.03583447 -0.06783842 -0.00126283  0.20052774  0.08933954  
 -0.06214588 -0.04138418  0.21468294 -0.21005437  0.0627281  0.03888206  
 -0.0815586  -0.14948195 -0.15470326  0.08970205  0.47115654  0.17413612  
 -0.14493531  0.03213331 -0.12277966 -0.0101483  0.1068163  0.01959883  
 -0.11612274  0.0174401  -0.07201762  0.01378155  0.13125205  0.08163541  
 -0.01263516  0.21001862  0.02932346 -0.00731285  0.04587183 -0.06869306  
 -0.10127032 -0.01449765 -0.06841056 -0.00676081  0.08469604 -0.11185972  
  0.06063811  0.09250596 -0.22186066  0.08084811 -0.02199327 -0.07330155  
  0.03902478  0.10345823 -0.14038587 -0.07197583  0.17452627 -0.19003527  
  0.1731386  0.15898383  0.03399535  0.15063891  0.03454215  0.09621807  
  0.02014965 -0.0044727 -0.1691166  -0.02025912  0.04263476 -0.06371888  
  0.03613986  0.01664142]
```

نمونه دوم Embedding



Hamidreza Encoding

```
[ -1.57438055e-01  4.51420657e-02  2.60103811e-02 -6.62469491e-02
-5.09227216e-02  4.10383269e-02 -2.05612928e-03 -3.41037959e-02
 2.65353560e-01 -1.29014120e-01  1.50531262e-01 -7.52204191e-03
-1.72026366e-01 -6.17446676e-02 -1.05310716e-02  1.25807256e-01
-1.99468285e-01 -1.08744271e-01 -7.48894811e-02 -1.01533242e-01
 8.55252221e-02  2.44798008e-02 -6.12826794e-02  1.06689297e-01
-2.16018125e-01 -3.73453051e-01 -1.82929486e-02 -9.81219262e-02
 3.07815745e-02 -7.78459162e-02 -8.24316591e-03  4.47502993e-02
-2.46524572e-01 -3.47807929e-02 -1.42760081e-02  1.34120494e-01
 2.45630108e-02  1.56474486e-02  1.23186693e-01  5.25238775e-02
-1.69254184e-01  1.68561488e-02 -2.08730791e-02  2.38551140e-01
 2.13326812e-01  7.97474235e-02 -2.06376370e-02 -3.99051979e-02
 1.81742653e-01 -2.56019205e-01  1.01454360e-02  1.35296300e-01
 2.79467702e-02  4.57503879e-03  7.69930407e-02 -1.66699380e-01
-4.07623034e-03  1.15864009e-01 -2.31608406e-01  9.44252461e-02
-2.39102282e-02 -1.43556297e-01 -1.08134327e-02  1.01711154e-02
 1.64106146e-01  4.50286344e-02 -9.14305449e-02 -1.22851834e-01
 1.88685060e-01 -1.70085356e-01  4.04173471e-02  5.06152548e-02
-5.88594861e-02 -2.18304336e-01 -3.05608988e-01  6.13882281e-02
 4.85496849e-01  1.81023002e-01 -1.52681708e-01  1.06852368e-01
-9.22335312e-02 -4.20196168e-02  1.18872024e-01  8.53119418e-02
-3.67120728e-02  5.07200975e-03 -7.40319118e-02  7.05566928e-02
 1.91654399e-01  1.00000896e-01 -1.63074266e-02  2.60287642e-01
-4.33973968e-02  2.68027354e-02  7.29569979e-03 -1.83562152e-02
-1.56751350e-01  1.25633413e-03 -8.51902887e-02 -7.04067126e-02
 3.18223312e-02 -4.51011583e-02  4.81905267e-02  7.68465996e-02
-2.26611674e-01  1.76405191e-01 -2.37509981e-02 -3.96564603e-04
-1.97020546e-02  3.96548323e-02 -1.43210605e-01  1.63690541e-02
 1.64528504e-01 -2.86139131e-01  1.84023544e-01  1.84423476e-01
 5.25022447e-02  2.08110526e-01  7.91555047e-02 -2.37734336e-03
 3.87546048e-02 -5.15421480e-03 -1.75652355e-01 -4.27460819e-02
 3.89414914e-02 -9.07934532e-02  1.30878732e-01  2.84349378e-02]
```


محاسبه شباهت چهره افراد

```
print("face_distance between gholamreza and hamidreza", face_recognition.face_distance([ghd_face_encoding], hrad_face_encoding)[0])
```

face_distance between gholamreza and hamidreza 0.6462456336677844

Euclidean distance = 0.6462...

این ۱۲۸ عدد چه قسمت هایی از صورت را معیار سنجی می کنند؟

○ ما هیچ نظری در این مورد نداریم. همه آن چیزی که برای ما اهمیت دارد این است که شبکه زمانی که به دو عکس مختلف از یک فرد مشخص نگاه می کند ، اعداد تقریباً یکسانی را تولید می کند.

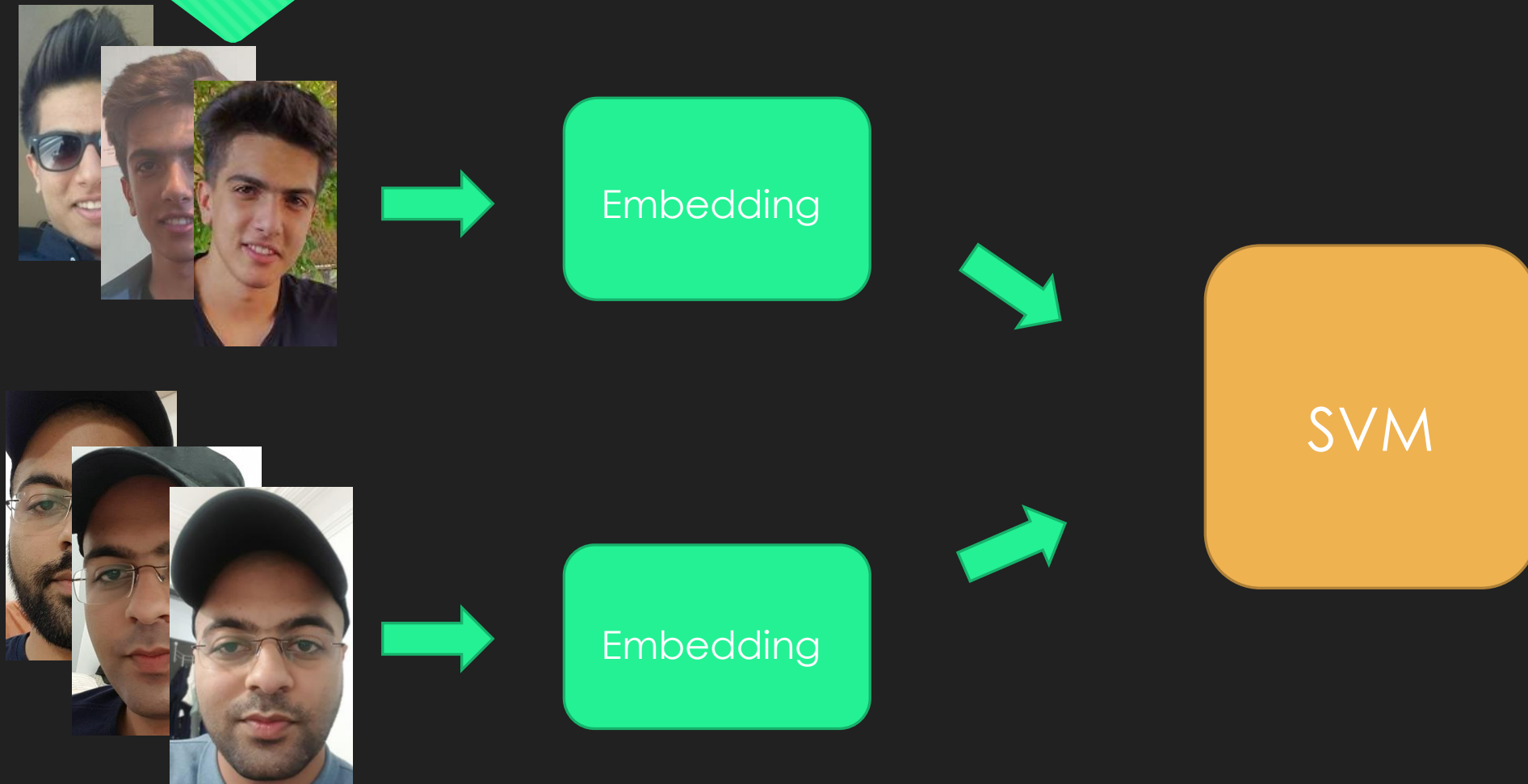
پیدا کردن نام فرد از روی رمز نگاشته شده

- شخصی را از پایگاه داده افراد شناخته شده که نزدیکترین معیار ها را به عکس آزمایشی ما دارد، پیدا کنیم.
- با استفاده از الگوریتم های کلاسه بندی (یک کلاسه بند خطی ساده به نام SVM)
- تمام کاری که باید انجام دهیم، آموزش یک کلاسه بند است که بتواند معیار ها را از یک تصویر آزمایشی جدید بگیرد و بگوید کدام فرد شناخته شده نزدیکترین تطابق را دارد. اجرای این کلاسه بند چند میلی ثانیه طول می کشد. نتیجه کلاسه بند، نام شخص مورد نظر می باشد!

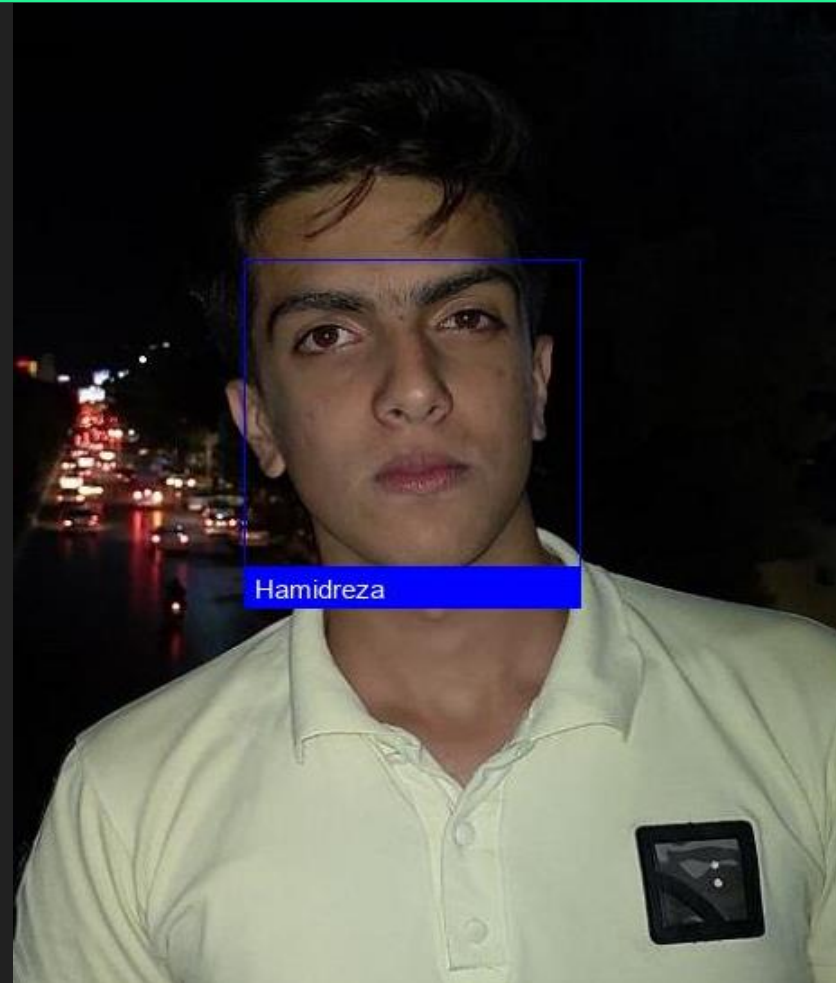
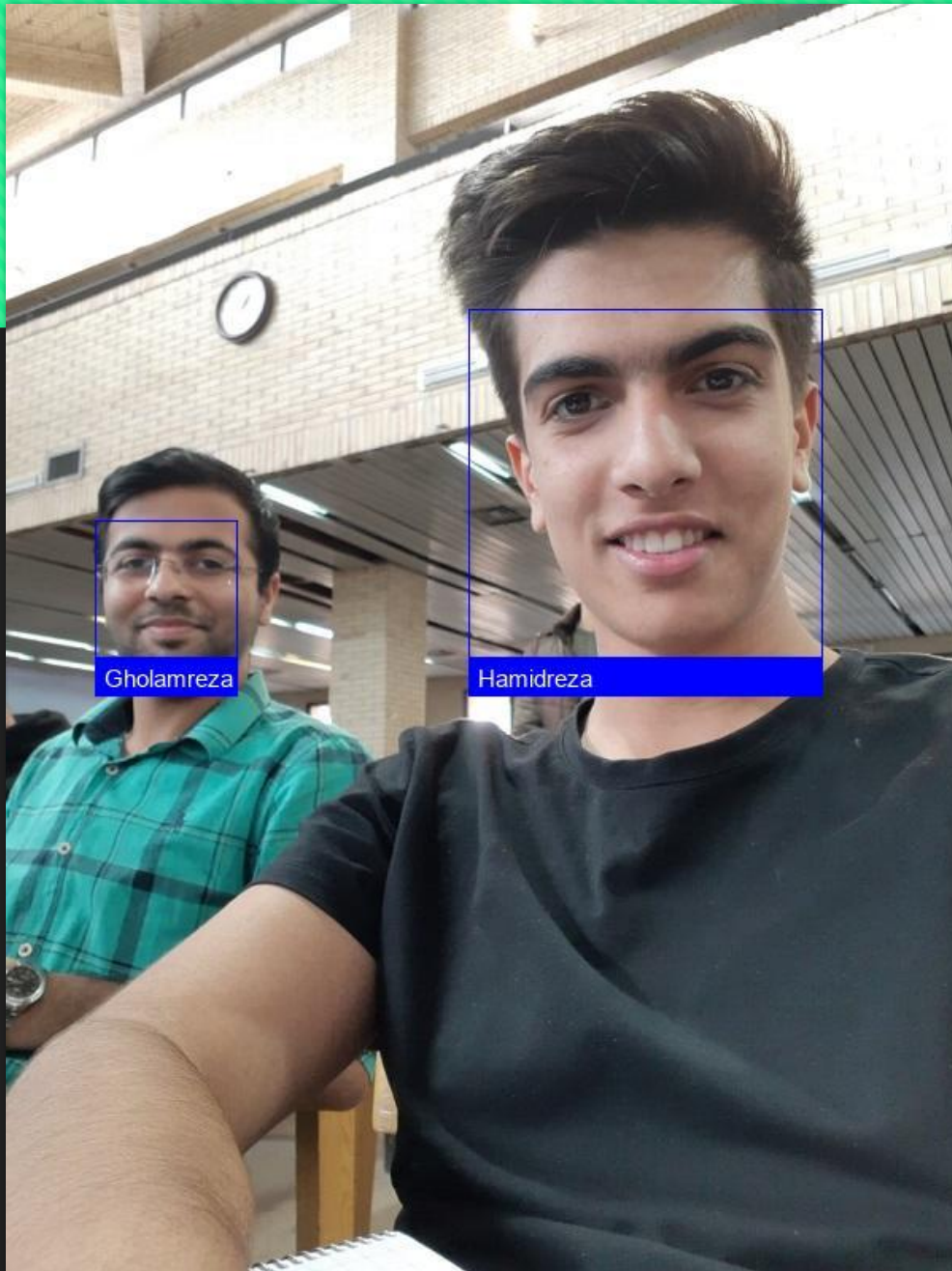
توضیح دمو ها

- تعدادی تصویر از افراد مورد نظر را به برنامه میدهیم
- برنامه تمام چهره ها را پیدا میکند
- برنامه Embedding چهره های موجود در تصاویر را محاسبه میکند
- Embedding ها را به همراه label مربوطه به svm جهت خوشه بندی میدهد

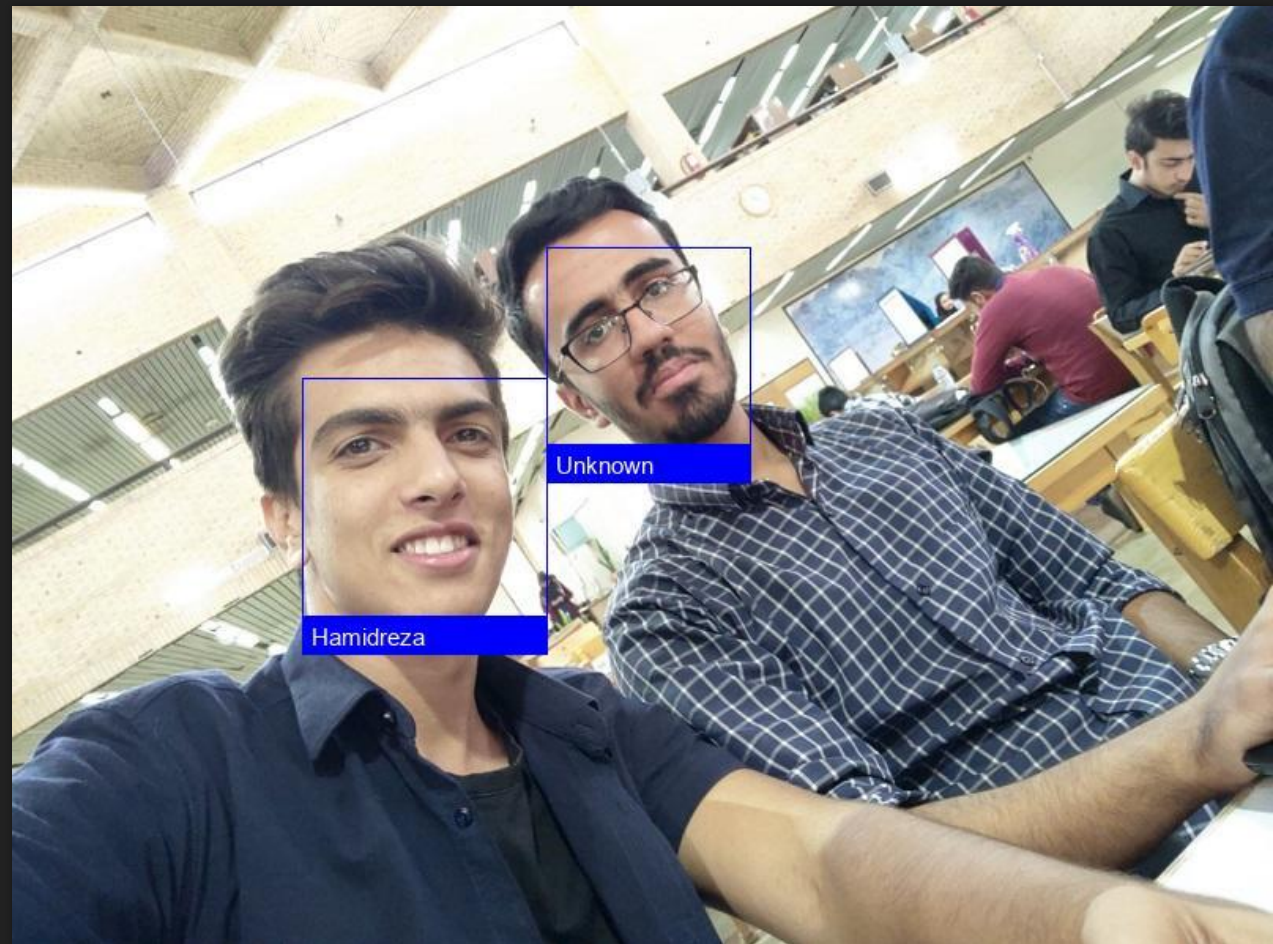
Training SVM



نتیجه ها



نتیجه ها (فرد ناشناس)



مصادر

- https://github.com/ageitgey/face_recognition
- <https://medium.com/@ageitgey/machine-learning-is-fun-part-4-modern-face-recognition-with-deep-learning-c3cffc121d78>

غلامرضا دار 96405037
حمیدرضا رادفر 96407013
علی سیف الدینی 96405047

دی 1399