



گزارش پروژه درس شناسایی الگو

امیرحسین امینی مهر 99722162

GoogleNet

بهار 1400

## فهرست مطالب:

3	مقدمه
4	مجموعه داده
6	پیش پردازش مجموعه داده
7	شبکه لنت
8	الکس نت
11	زد اف نت
12	وی جی جی
14	گوگل نت
15	دسته بند های کمکی
15	ماژول آغازین
20	نتیجه گیری و مقایسه
20	بهترین دقت روی داده های تست
20	نمودار تغییرات دقت
24	نمودار تغییرات خروجی تابع ضرر
27	ماتریس درهم ریختگی
32	دقت سه کلاس برتر

## مقدمه:

یکی از چالش‌های موجود در یادگیری عمیق انتخاب یک معماری یا مدل مناسب برای رسیدن به دقت مناسب در یک تسک می‌باشد که همین امر باعث شده که تحقیقات بسیاری روی معماری‌های مختلف در یادگیری عمیق انجام شود و معماری‌های مختلفی به وجود بیایند.

یکی از تسک‌های مهمی که تحقیقات زیادی روی آن انجام شده است دسته‌بندی<sup>1</sup> بندی تصاویر مختلف می‌باشد به این صورت که تعدادی تصویر و تعدادی برچسب<sup>2</sup> داریم که هر کدام از این تصاویر نماینده فقط یکی از برچسب‌ها می‌باشند. (به عنوان مثال یکی از تسک‌های معروف در این زمینه تشخیص اعداد در مجموعه داده<sup>3</sup> Mnist می‌باشد که در واقع این مجموعه داده شامل تصاویر اعداد دست نویس بین یک تا نه می‌باشند که مدل ما باید تشخیص دهد که هر تصویر چه عددی را نشان می‌دهد)

به دلیل اهمیت اینگونه تسک‌ها یک مسابقه تحت عنوان چالش تشخیص تصویری مجموعه داده ایمیج نت<sup>4</sup> شکل گرفت که در این مسابقه که تا سال 2017 ادامه داشت شرکت کنندگان باید یک معماری برای حل تسک دسته‌بندی مجموعه داده ایمیج نت<sup>5</sup> (در واقع ایمیج نت یک مجموعه داده دارای حدود 16 میلیون تصویر مختلف از 1000 برچسب مختلف می‌باشد که هر یک تصاویر فقط به یک برچسب تعلق دارند و تصاویر مربوط به اشیاء و خوراکی‌ها و حیوانات و... می‌باشند) معرفی می‌کردند و هر معماری که دقت بیشتری داشت برنده مسابقه می‌شد و این مسابقه تا چند سال ادامه داشت و هر سال یک معماری که دقت بیشتری داشت برنده می‌شد تا اینکه در سال 2017 به دلیل اینکه دقت معماری‌های ارائه شده از دقت انسان هم به میزان خوبی بیشتر شد مسابقه به اتمام رسید و دیگر ادامه پیدا نکرد.

اولین معماری‌ای که توانست از دقت انسان در دسته‌بندی مجموعه داده ایمیج نت پیشی بگیرد گوگل نت<sup>6</sup> بود که نسبت به معماری‌های قبلی تفاوت زیادی داشت و هدف اصلی این پروژه بررسی این معماری می‌باشد ولی برای درک بهتر قدرت این معماری چهار معماری دیگر که قبل از گوگل نت معرفی شده بودند نیز در پروژه پیاده‌سازی شده است و با معماری گوگل نت مقایسه شده اند که این معماری‌ها لنت<sup>7</sup>، الکس نت<sup>8</sup>، زد اف نت<sup>9</sup> و وی جی جی<sup>10</sup> 16 می‌باشند که در ادامه هر پنج معماری به طور کامل بررسی خواهند شد و نتایج بدست آمده در پروژه از مقایسه اعمال این معماری‌ها روی مجموعه داده تشخیص گل<sup>11</sup> گزارش خواهد شد.

---

<sup>1</sup> Classification

<sup>2</sup> Label

<sup>3</sup> Dataset

<sup>4</sup> ImageNet Large Scale Visual Recognition Challenge

<sup>5</sup> ImageNet

<sup>6</sup> GoogleNet

<sup>7</sup> LeNet

<sup>8</sup> AlexNet

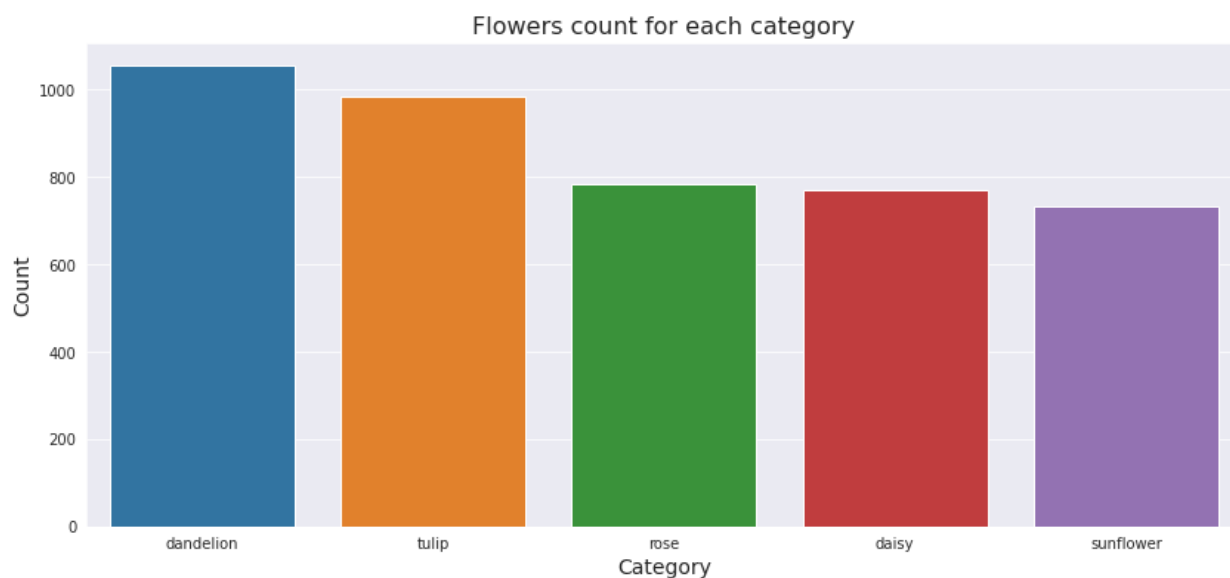
<sup>9</sup> ZFNet

<sup>10</sup> VGG-16

<sup>11</sup> Flowers-Recognition

## مجموعه داده:

در این پروژه از مجموعه داده تشخیص گل که بالاتر به آن اشاره شد استفاده شده است که از 4242 تا تصویر گل مختلف تشکیل شده است که این گل ها در پنج کلاس دسته بندی می شوند که برچسب های آن گل های مروارید<sup>12</sup>، قاصدک<sup>13</sup>، رز<sup>14</sup>، آفتاب گردان<sup>15</sup> و گل لاله<sup>16</sup> می باشند. تعداد تصاویر موجود از هر گل در مجموعه داده به شکل ذیل می باشد.



نمودار 1- تعداد تصاویر موجود از هر گل در مجموعه داده تشخیص گل

---

<sup>12</sup> Daisy

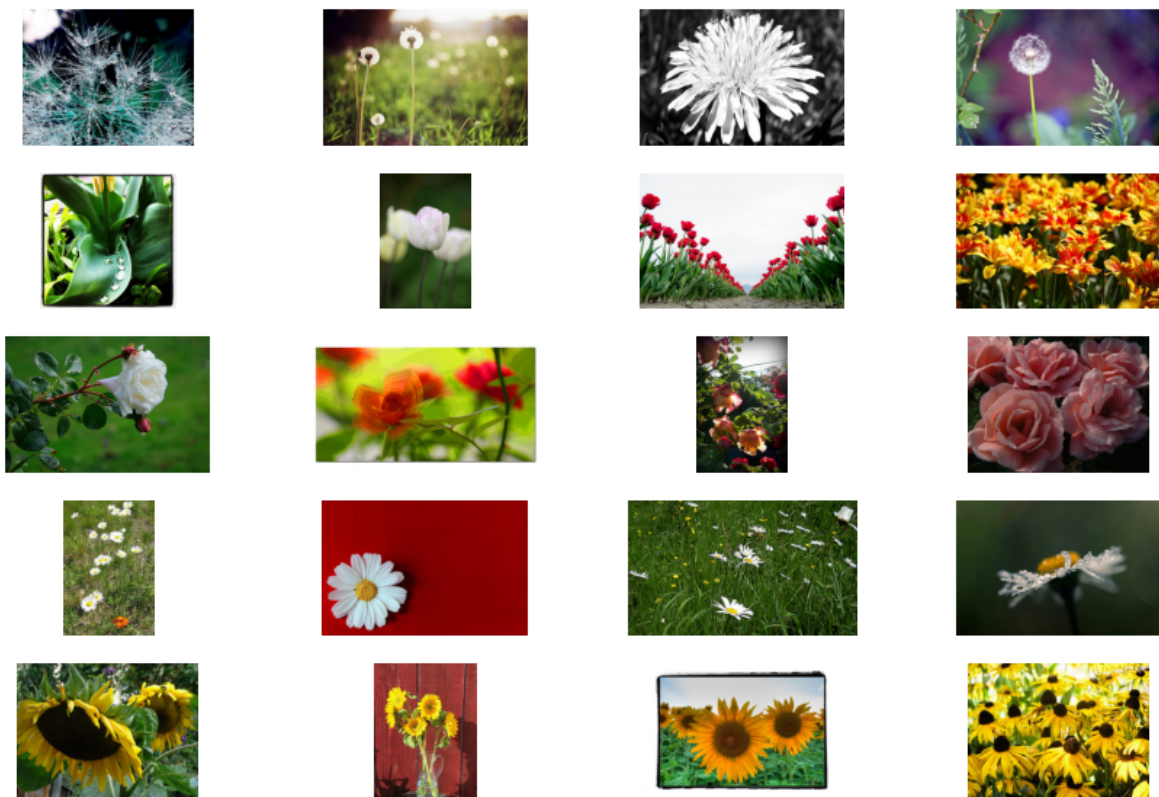
<sup>13</sup> Dandelion

<sup>14</sup> Rose

<sup>15</sup> Sunflower

<sup>16</sup> Tulip

نمونه هایی از تصاویر موجود در این مجموعه داده نیز به شکل ذیل می باشد.

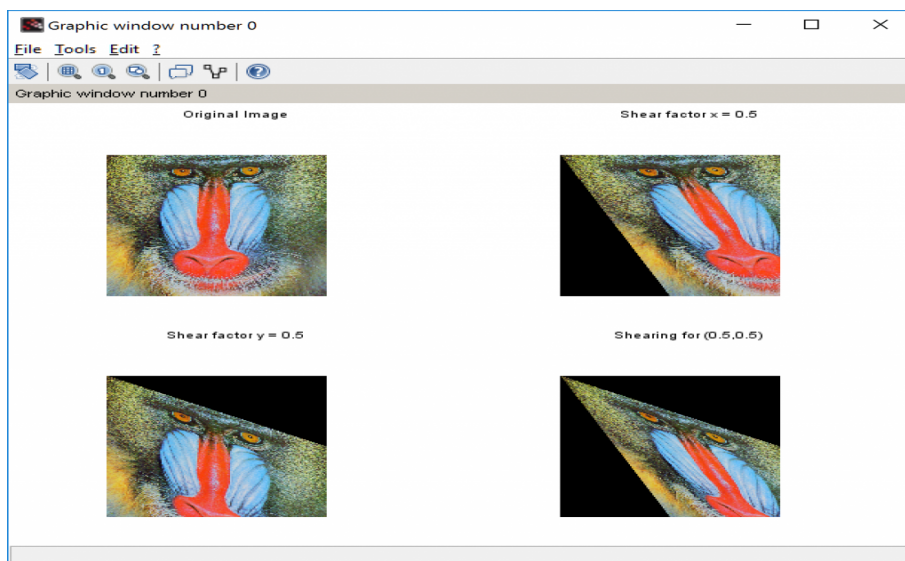


شکل 1- نمونه هایی از تصاویر موجود در مجموعه داده تشخیص گل

## پیش پردازش مجموعه داده:

در پروژه برای استفاده از مجموعه داده تشخیص گل باید ابتدا آن را به دو بخش داده های آموزشی و داده های تست تقسیم کنیم زیرا وقتی مجموعه داده را دانلود می کنیم دارای پنج فولدر می باشد که هر فولدر مربوط به تصاویر گلی هست که نام آن فولدر می باشد پس برای اینکار باید در ابتدا دو فولدر به نام های تست و ترین تشکیل داد و درون هر کدام پنج فولد برای هر گل به وجود آورد و از عکس های موجود در هر فولدر از پنج فولدر ابتدایی صد تا را در فولدر های مربوط به داده های تست قرار داد و مابقی را در فولدر مربوط به داده های آموزشی پس در نهایت دو فولدر خواهیم داشت که هر کدام پنج فولدر درون خود دارند که این پنج فولدر بر اساس نام برجسب ها که قبل تر به اشاره شد نام گذاری شده اند و حال شرایط محیا می باشد که مولد<sup>17</sup> های تست و ترین به وجود بیایند که در مولد مربوط به داده های آموزشی پیش پردازش های ذیل را خواهیم داشت:

- نرمال سازی تصاویر با تقسیم کردن مقدار روشنایی پیکسل های هر تصویر بر عدد 255 که در واقع حداکثر میزان روشنایی هر پیکسل می باشد.
- داده افزایی<sup>18</sup> با استفاده از شیفیت تصاویر به اندازه صفر تا بیست درصد در عرض.
- داده افزایی با استفاده از شیفیت تصاویر به اندازه صفر تا بیست درصد در طول.
- داده افزایی با استفاده از برش<sup>19</sup> بین صفر تا بیست درصد از تصاویر (در واقع برای اینکار تصویر به نوعی جمع می شود و قسمت های باقی مانده سیاه می شوند) که برای فهم بهتر این قسمت یک تصویر نمونه در ذیل آورده شده است.



شکل 2- نمونه ای از برش تصویر

<sup>17</sup> Generator

<sup>18</sup> Data Augmentation

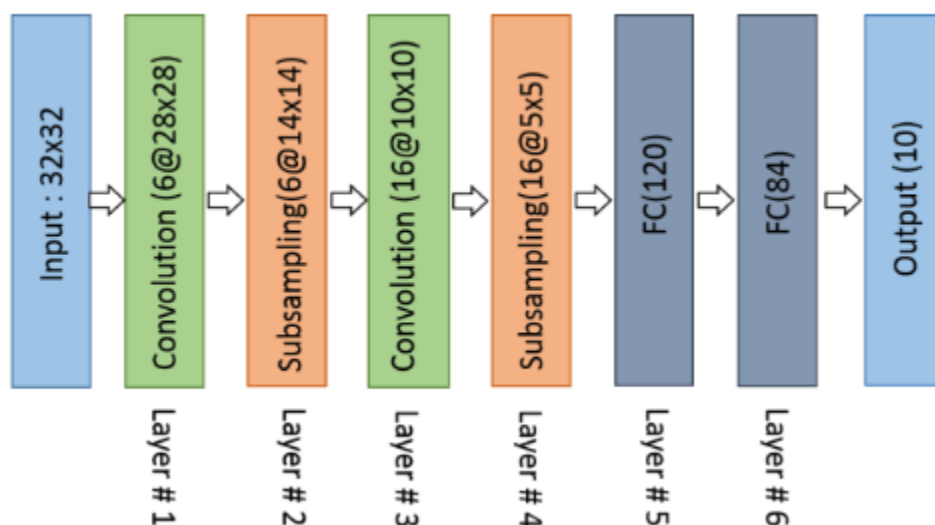
<sup>19</sup> Shear

- داده افزایی با استفاده از صفر تا بیست درصد زوم کردن در تصاویر.
- داده افزایی با استفاده از وارونه کردن تصاویر به صورت افقی.

در مدل مولد مربوط به داده های تست هم فقط تصاویر نرمالیزه شده اند و مقدار روشنایی هر پیکسل بر عدد 255 تقسیم شده است. حال باید با استفاده از مولد ها مدل ها را آموزش داد و دقت آن ها بر روی داده های تست را بدست آورد.

### شبکه لنت:

اولین شبکه ی کانولوشنی که بسیار مطرح شد شبکه ی لنت بود که در مقاله ی نام برده شده در پاورقی<sup>20</sup> اولین بار معرفی شد و نویسنده آن لکون<sup>21</sup> بود و در سال 1998 ارائه شد. نام شبکه نیز از نام لکون نویسنده این مقاله برداشته شده است. این شبکه را می توان به عنوان اولین شبکه که فیلتر های کانولوشنی در آن استفاده شده است نام برد. هدف این شبکه، خواندن کد پستی از پشت بسته های پستی بود.



شکل 3- لایه های شبکه لنت

شکل بالا شبکه ی لنت را نمایش می دهد. این شبکه در ابتدا یک ورودی  $32 \times 32$  دریافت می کند و سپس از 6 فیلتر کانولوشنی  $5 \times 5$  که گام<sup>22</sup> آن برابر یک است عبور می کند. حاصل یک تانسور<sup>23</sup>

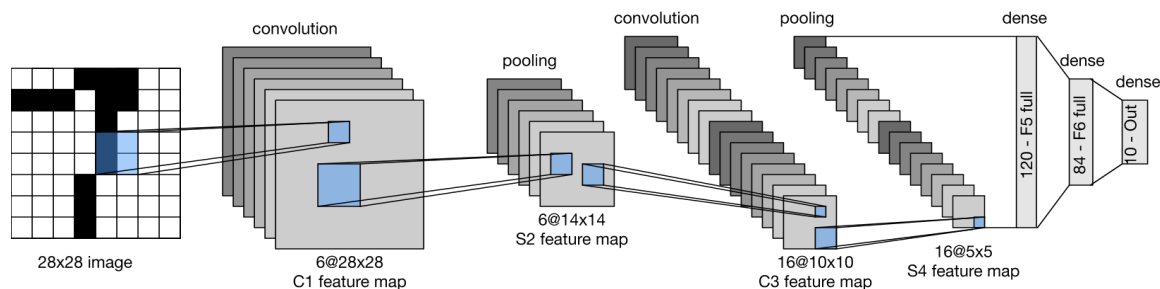
<sup>20</sup>Gradient-based learning applied to document recognition

<sup>21</sup> LeCun

<sup>22</sup> Stride

<sup>23</sup> Tensor

$28 \times 28 \times 6$  است. سپس از یک فیلتر زیر نمونه گیری<sup>24</sup> که در این جا یک فیلتر  $2 \times 2$  تجميع میانگین<sup>25</sup> با گام دو است عبور می کند. این فیلتر موجب می شود که طول و عرض آرایه نصف شود اما عمق آن تغییری نمی کند در نتیجه به یک آرایه  $14 \times 14 \times 6$  می رسیم. این بار از 16 فیلتر کانولوشنی  $5 \times 5$  با گام یک عبور می دهیم که حاصل  $10 \times 10 \times 16$  می شود. دوباره در ادامه از یک لایه تجميع میانگین مانند قبل عبور می دهیم که حاصل  $5 \times 5 \times 16$  می شود. و در انتها نیز از دو لایه کاملاً متصل<sup>26</sup> عبور داده می شوند و شکل کلی به صورت ذیل خواهد بود.



شکل 4- شکل کلی شبکه لنت

## الکس نت:

الکس نت دومین شبکه کانولوشنی شناخته شده می باشد که در سال 2012 و با فاصله چهارده ساله با لنت و با پیشرفت زیاد سخت افزار کامپیوتر ها معرفی شد. این شبکه که برای اولین بار نوآوری های زیادی داشت دارای هشت لایه (پنج لایه ی کانولوشنی و سه لایه کاملاً متصل) است که موجب شده در دسته ی شبکه های کم عمق<sup>27</sup> قرار گیرد. از نوآوری های این معماری می تواند به معرفی تابع یکسوساز<sup>28</sup> که امروزه یکی از پرکاربردترین تابع های فعال سازی می باشد و معرفی لایه حذف تصادفی<sup>29</sup> که امروزه یکی از لایه های پرکاربرد برای رگولاریزیشن<sup>30</sup> می باشد اشاره کرد. الکس نت در مسابقه تشخیص تصویری مجموعه داده ایمج نت در سال 2012 با کسب خطای 5 کلاس برتر معادل با 15.4 درصد رتبه اول را کسب کرد. معماری کلی این شبکه را می توان در تصویر ذیل مشاهده کرد.

<sup>24</sup> Subsampling

<sup>25</sup> AveragePool

<sup>26</sup> Fully connected

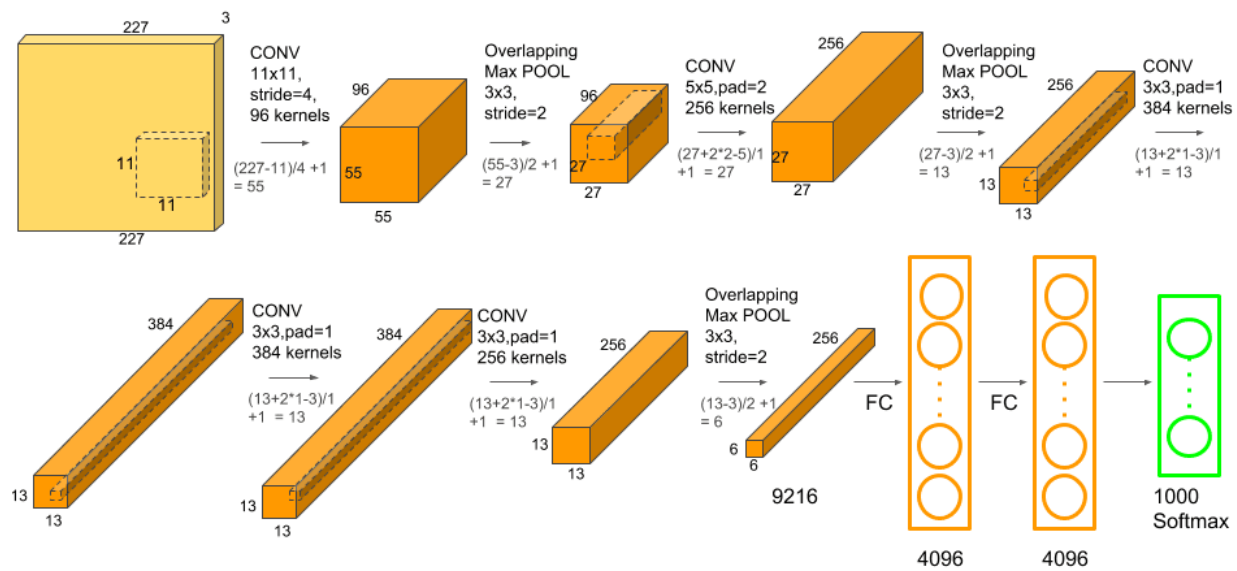
<sup>27</sup> Shallow

<sup>28</sup> ReLU

<sup>29</sup> Dropout

<sup>30</sup> Regularization





شکل 5- شکل کلی شبکه الکس نت

شکل بالا شبکه ی الکس نت را نمایش می دهد. (دلیل این که لایه ها دوطبقه ای هستند، استفاده از دو پردازنده گرافیکی<sup>31</sup> برای پردازش آن است) در شکل ورودی شبکه یک عکس  $227 \times 227$  است در ابتدا 96 فیلتر کانولوشنی  $11 \times 11$  همراه با چهار گام اعمال می شود. خروجی یک آرایه ی  $55 \times 55 \times 96$  است. (تعداد پارامتر های هر لایه ی کانولوشنی برابر است با: طول لایه \* عرض لایه \* عمق ورودی \* تعداد فیلتر ها (به طور مثال در اینجا برابر است با  $96 \times 3 \times 11 \times 11$ )) در مرحله ی بعد از یک لایه ی  $3 \times 3$  تجمیع حداکثر<sup>32</sup> با اندازه ی دو گام عبور داده می شود که خروجی  $27 \times 27 \times 96$  می شود. (تعداد پارامتر های قابل یادگیری در تجمیع حداکثر صفر است). در مرحله بعد از 256 فیلتر کانولوشنی  $5 \times 5$  با گام یک و لایه گذار<sup>33</sup> دو عبور می دهیم که خروجی  $27 \times 27 \times 256$  می باشد. در مرحله ی بعد نیز دوباره از فیلتر تجمیع حداکثر  $3 \times 3$  با گام دو عبور می دهیم که خروجی  $13 \times 13 \times 96$  می باشد. در سه مرحله ی بعد از فیلتر های کانولوشنی با گام و لایه گذار یک استفاده می کنیم که تعداد آن ها به ترتیب 384 ، 384 ، 256 است که خروجی نهایتاً  $13 \times 13 \times 256$  می شود. سپس از یک لایه ی تجمیع حداکثر  $3 \times 3$  با گام یک عبور می دهیم و خروجی آن  $6 \times 6 \times 256$  می شود. در آخر نیز سه لایه ی کاملاً متصل قرار دارد. دو لایه 4096 تایی و یک لایه 1000 تایی با تابع فعالسازی سافت مکس<sup>34</sup> برای تعیین کلاس های موجود در مساله. (به طور مثال اگر قصد انتخاب بین سه کلاس سگ، گربه، موش را داشته باشیم اندازه لایه ی آخر را سه در نظر می گیریم).

<sup>31</sup> GPU

<sup>32</sup> Maxpool

<sup>33</sup> Padding

<sup>34</sup> Softmax

در تصویر ذیل می توان تفاوت معماری های لنت و الکس نت را مشاهده کرد که همانطور که مشخص است الکس نت عمیق تر از لنت می باشد و قابلیت یادگیری بیشتری دارد.

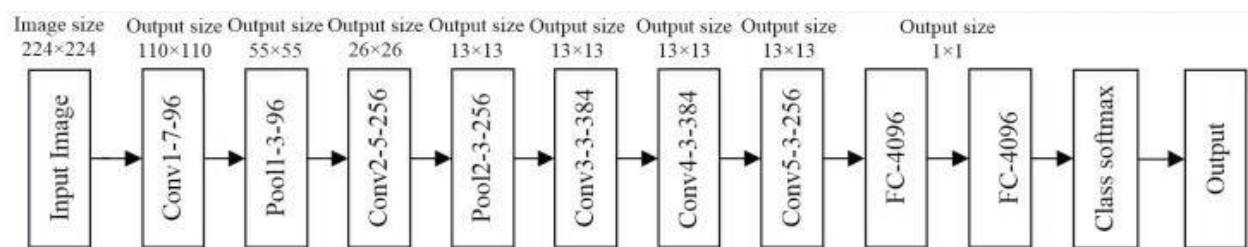


شکل 6- مقایسه لایه های شبکه های لنت و الکس نت

## زد اف نت:

زد اف سومین شبکه کانولوشنی شناخته شده می باشد که در سال 2013 توسط راب فرگوس<sup>35</sup> و متیو زیلر<sup>36</sup> معرفی شد که در واقع نام این معماری هم از نام این دو فرد گرفته شده است. زد اف نت در مسابقه تشخیص تصویری مجموعه داده ایمچ نت در سال 2013 با کسب خطای 5 کلاس برتر معادل با 14.8 درصد رتبه اول را کسب کرد.

این معماری نسبت به الکس نت نوآوری خاصی ندارد و مهم ترین نوآوری آن قسمت تجسم<sup>37</sup> آن است. ابزاری که با استفاده از آن، کاربرد هر نورون در شبکه تا حدی مشخص می شود و به درک شبکه های کانولوشنی کمک بزرگی می کند. در این ابزار می توان دید که در لایه های نخست، شبکه خطوط و لبه های ساده را تشخیص می دهد و در لایه های بعد نورون ها مفاهیم پیچیده تری مانند صورت انسان یا نوشته را تشخیص می دهند. بعضی از نورون ها برای ما قابل فهم نیستند. ولی معلوم شده حذف آن ها هم تاثیری مانند حذف دیگر نورون ها در عملکرد نهایی شبکه دارد. لایه های این شبکه به صورت ذیل می باشند.



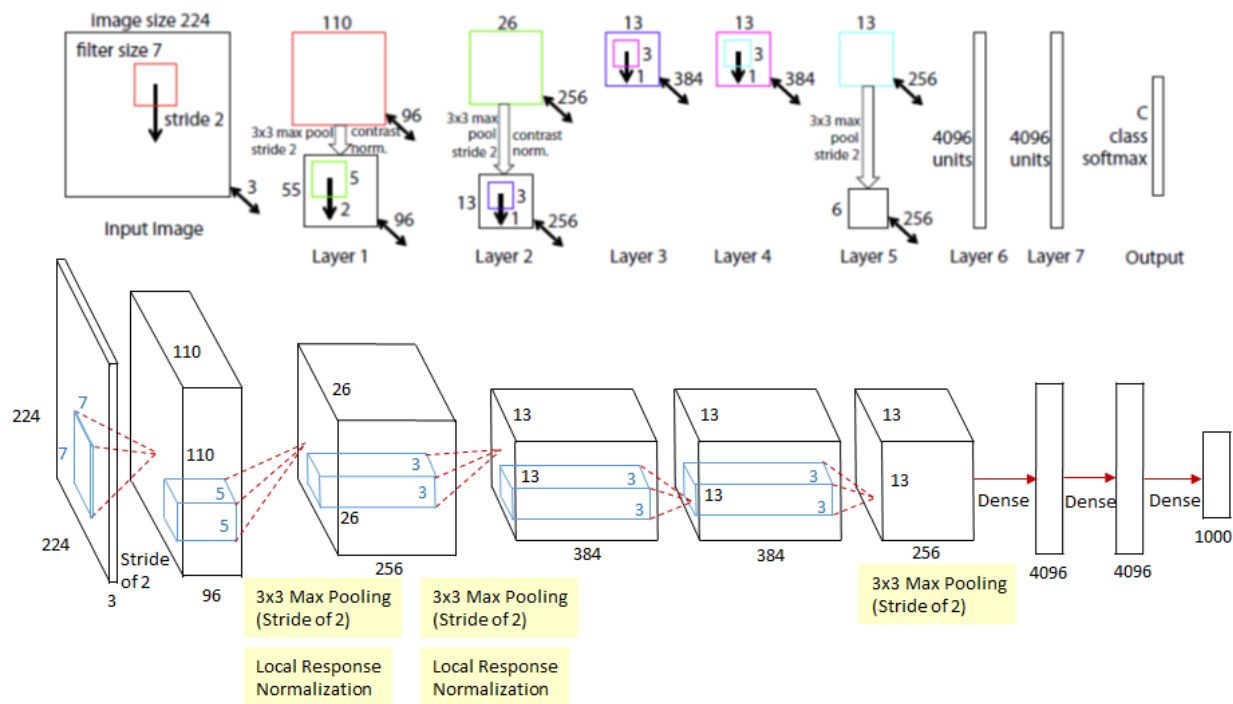
شکل 7- لایه های شبکه زد اف نت

همانطور که در شکل بالا مشخص می باشد این شبکه تا حد زیادی همان معماری شبکه الکس نت را حفظ کرده است با این تفاوت که در لایه اول، به جای یک لایه کانولوشن با سایز 11\*11 همراه با چهار گام از یک لایه کانولوشن با سایز 7\*7 همراه با چهار گام استفاده شده است و دلیل آن هم این بود که در الکس نت سایز فیلتر بزرگ گرفته شده بود با گام زیاد که این کار بیهوده بود و همان کار را می شد با یک لایه کانولوشن با سایز کوچکتر و گام کمتر انجام داد. تفاوت دیگر آن با الکس نت این است که در لایه های 3، 4 و 5 به ترتیب به جای 384، 384 و 256 فیلتر، از 512، 1024 و 512 فیلتر استفاده شده است. تصویر ذیل شکل کلی این معماری را نشان می دهد.

<sup>35</sup> Rob Fergus

<sup>36</sup> Matthew Zeiler

<sup>37</sup> Visualization



شکل 8- شکل کلی شبکه زد اف نت

## وی جی جی:

وی جی جی چهارمین شبکه کانولوشنی شناخته شده می باشد که در سال 2014 توسط سیمونیان<sup>38</sup> و زیسرمن<sup>39</sup> از دانشگاه آکسفورد در مقاله گفته شده در پاورقی<sup>40</sup> معرفی شد. با توجه به پیشرفت های اخیر شبکه های عصبی عمیق، شاید دیگر نتوان شبکه وی جی جی را یک شبکه مدرن دانست، اما به دلیل ساختار خوب، سادگی، تعداد لایه های نه چندان زیاد هنوز در بسیاری از زمینه های بینایی کامپیوتر از این شبکه استفاده می شود. بعضی از محققان ممکن است شبکه از پیش آموزش دیده وی جی جی را برای کاربرد مدنظر خود بازتنظیم<sup>41</sup> کنند و یا اینکه تغییراتی در معماری این شبکه محبوب ایجاد نمایند و از ابتدا<sup>42</sup> آن را آموزش دهند. در هر صورت، چه برای افراد حرفه ای در زمینه شبکه های عصبی و یادگیری ماشین و چه برای افراد تازه وارد، شناخت معماری شبکه وی جی جی بسیار کارآمد خواهد بود. در ادامه در مورد معماری شبکه عصبی عمیق وی جی جی توضیح داده خواهد شد. لایه های شبکه وی جی جی 16 به شکل ذیل می باشد.

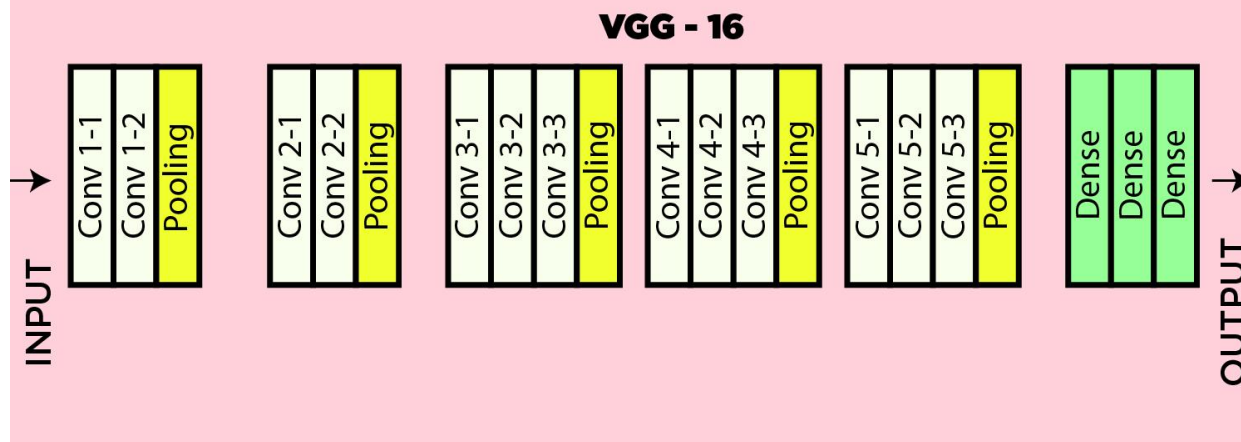
<sup>38</sup> K. Simonyan

<sup>39</sup> A. Zisserman

<sup>40</sup> Very Deep Convolutional Networks for Large-Scale Image Recognition

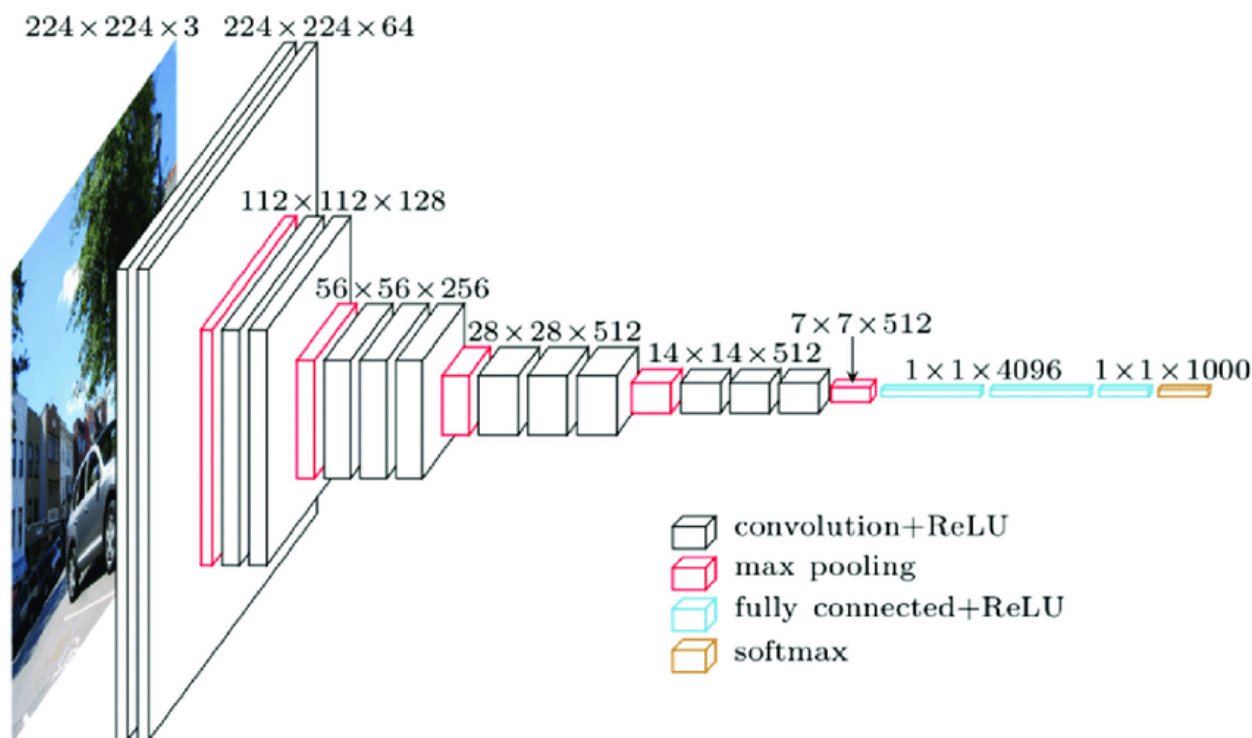
<sup>41</sup> Fine Tune

<sup>42</sup> Scratch



شکل 9- لایه های شبکه وی جی جی 16

شبکه وی جی جی در دو معماری مختلف با عنوان های وی جی جی 16 و وی جی جی 19 ارائه شده است. ابتدا شبکه های وی جی جی 16 پیشنهاد شد و بعدها با تغییراتی جزئی در شبکه وی جی جی 16 ، شبکه وی جی جی 19 مطرح گردید. شبکه وی جی جی 16 همان طور که در شکل بالا نشان داده شده، شامل 16 لایه کانولوشنی یا 16 لایه پارامتری است. شبکه وی جی جی 16 ، شامل دو لایه کانولوشنی با 64 فیلتر  $3 \times 3$  هست که پشت سر هم قرار گرفته اند. سپس، یک لایه ماکس پولینگ  $2 \times 2$  با گام به اندازه دو قرار گرفته است. این لایه ماکس پولینگ علاوه بر نمونه برداری، وظیفه کاهش بعد ویژگی ها به نصف را هم دارد. در ادامه، دو لایه کانولوشنی دیگر با 128 فیلتر  $3 \times 3$  و یک لایه ماکس پولینگ  $2 \times 2$  و پرش 2 قرار گرفته اند. به طور مشابه، سه لایه کانولوشنی با 256 فیلتر  $3 \times 3$  و یک لایه ماکس پولینگ  $2 \times 2$  با پرش 2 قرار گرفته اند. 3 لایه کانولوشنی با 512 فیلتر  $3 \times 3$  و یک لایه ماکس پولینگ در ادامه این شبکه قرار دارد که البته دو بار تکرار می شود. در نهایت، ویژگی ها تبدیل به یک بردار ویژگی می شوند تا در اختیار لایه های نورونی یا کاملاً متصل قرار گیرند. دو لایه نورونی به ابعاد 4096 پشت سر هم قرار گرفته اند. در نهایت، یک لایه نورونی به ابعاد 1000 که متناظر با تعداد کلاس های کاربرد ما هست، در نظر گرفته شده است. با توجه به اینکه پایگاه داده ایمج نت شامل 1000 کلاس هست، در اینجا هم لایه خروجی شامل 1000 نرون است. در تمامی لایه های کانولوشنی و لایه های نورونی از تابع فعال سازی یکسوساز استفاده شده است. در شکل ذیل می توان معماری کلی شبکه وی جی جی را مشاهده کرد.



شکل 10- شکل کلی شبکه وی جی جی 16

این معماری بر عکس معماری های قبلی نتوانست رتبه اول را در مسابقه تشخیص تصویری مجموعه داده ایمیج نت در سال 2014 کسب کند و با کسب خطای 5 کلاس برتر معادل با 7.3 درصد رتبه دوم را کسب کرد و گوگل نت توانست رتبه اول را در آن سال کسب کند که در ادامه به طور کامل گوگل نت مورد بررسی قرار خواهد گرفت.

## گوگل نت:

این شبکه توسط شرکت گوگل<sup>43</sup> در سال 2014 ارائه شد. نام آن ادای احترامی به یان لکون<sup>44</sup>، سازنده لنت، اولین شبکه کانولوشنی، است. این شبکه با ایده عمیق تر کردن شبکه های کانولوشنی به وجود آمده است. گوگل نت که دارای 22 لایه است با ارائه قویترین مدل در سال 2014، برنده مسابقه تشخیص تصویری مجموعه داده ایمیج نت شد.

در شبکه های قبلی معرفی شده یک مشکل بزرگ وجود داشت که اگر لایه های شبکه افزایش می یافت و تعداد لایه ها خیلی زیاد می شد محو شدگی گرادینان<sup>45</sup> رخ می داد به این معنا که وقتی تعداد لایه های شبکه خیلی زیاد می شد از آنجایی که در روش پس انتشار<sup>46</sup> برای اصلاح وزن های هر شبکه نیاز

<sup>43</sup> Google

<sup>44</sup> Yann Lecun

<sup>45</sup> Gradient Vanishing

<sup>46</sup> Back Propagation

است که از قاعده زنجیره ای<sup>47</sup> مشتق استفاده شود و در قاعده زنجیره ای به ازای هر لایه یک مشتق باید حساب شود و ضرب در باقی مشتق های مربوط به لایه های جلوتر شود اگر مقدار یک مشتق در لایه های جلوتر از یک لایه صفر یا نزدیک صفر شود عملاً ضرب حاصل را کوچک کرده و وزن های آن لایه نمی توانند اصلاح شوند و هر چقدر تعداد لایه ها زیاد تر شود احتمال رخ دادن این مساله زیاد تر می شود و آنقدری تعداد ضرب ها زیاد می شود که هیچ فیدبکی از مقدار تابع ضرر به لایه های ابتدایی نمی رسد و نمی توانند اصلاح شوند که سازندگان گوگل نت برای حل این مشکل از دسته بند های کمی<sup>48</sup> استفاده می کند که در ادامه بررسی خواهد شد.

### دسته بند های کمی:

دسته بند های کمی در واقع برای حل مشکل محو شدگی گرادیان به وجود آمده اند و برای اینکار در گوگل نت دو دسته بند کمی وجود دارد که از ویژگی های استخراج شده در لایه های ابتدایی و میانی برای دسته بندی استفاده می کند و تابع ضرر<sup>49</sup> نهایی مجموع مقادیر خروجی تابع ضرر دو دسته بند کمی و تابع ضرر کلی شبکه می باشد و در واقع دو دسته بند های کمی چون از لایه های ابتدایی و میانی زیاد دو نیستند می توانند مقادیر خروجی تابع ضرر خود را به خوبی پس انتشار کنند و با این روش وزن های لایه های ابتدایی و میانی شبکه هم خواهند توانست به خوبی اصلاح شوند و دچار محو شدگی گرادیان نخواند شد.

اما مهم ترین نوآوری گوگل نت مورد دیگری به نام ماژول آغازین<sup>50</sup> می باشد که در که در معماری های بعد از آن نیز بسیار از آن استفاده شده است و امروزه یکی از پرکاربرد ترین روش های مورد استفاده در این زمینه می باشد که در ادامه بررسی خواهد شد.

### ماژول آغازین:

در این ماژول، ۳ فیلتر کانولوشنی با اندازه های متفاوت و یک فیلتر پولینگ روی لایه قبلی اجرا شده و نتایج را به صورت یک تنسور واحد کنار هم می گذارد. در این صورت شبکه می تواند در لایه بعدی تصمیم بگیرد که از نتیجه کدام فیلتر و به چه اندازه می خواهد استفاده کند. ساینز فیلتر های کانولوشنی به صورتی که در تصویر مشخص است،  $1*1$ ،  $3*3$  و  $5*5$  می باشد و فیلتر پولینگ با ساینز  $3*3$  روی آن اجرا شده است.

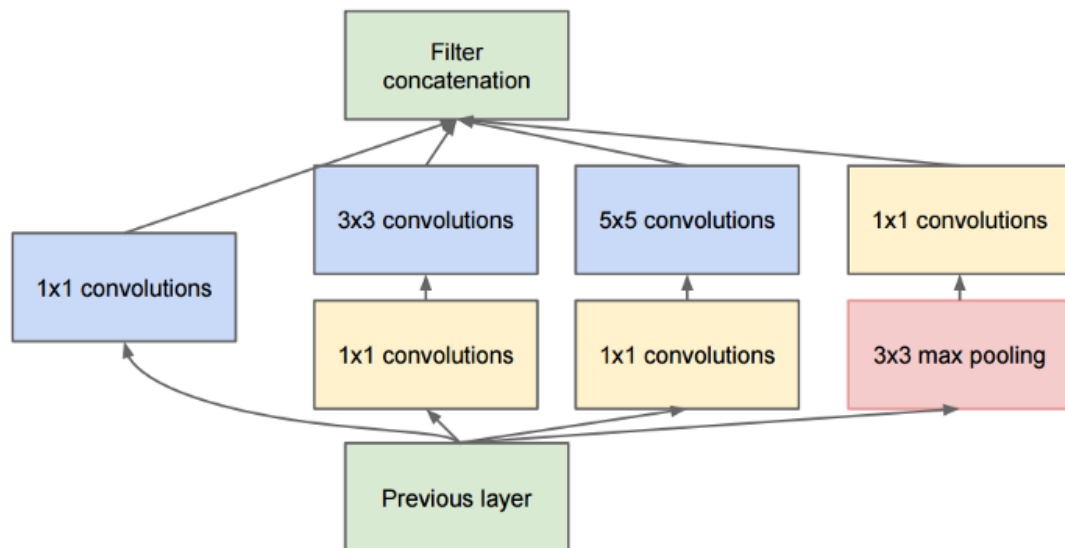
---

<sup>47</sup> Chain Rule

<sup>48</sup> Auxiliary Classifier

<sup>49</sup> Loss Function

<sup>50</sup> Inception module



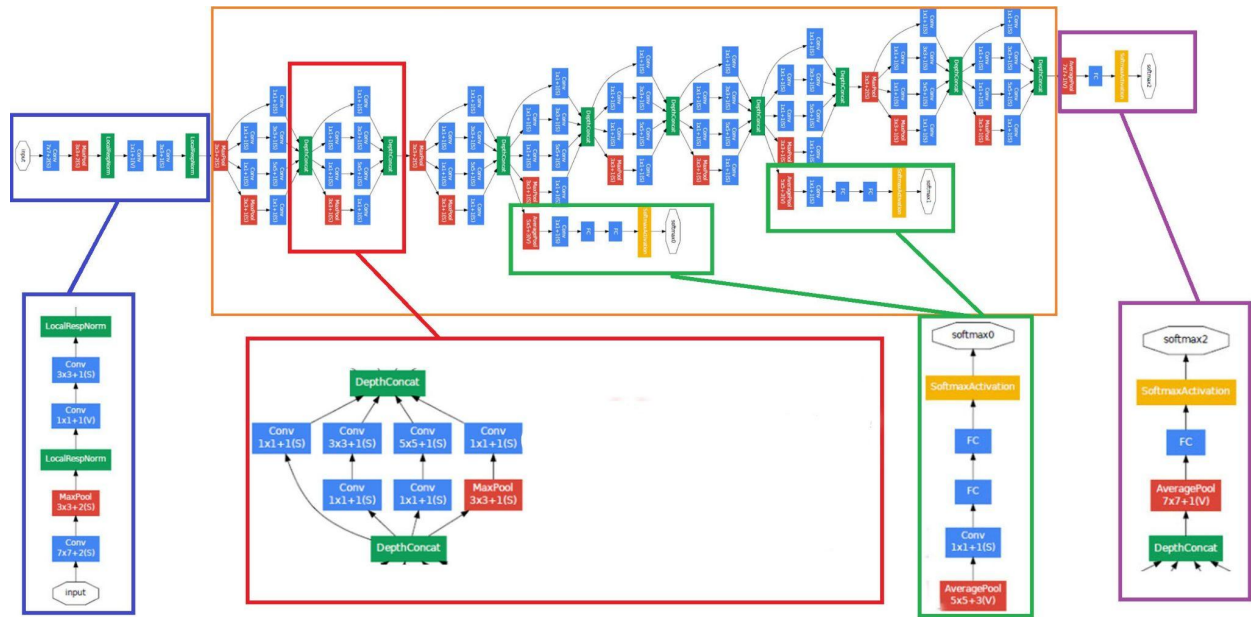
شکل 11- شکل کلی ماژول آغازین

همانطور که در تصویر بالا می‌بینید، این ماژول یک تغییر بزرگ نسبت به معماری‌های قبلی بررسی شده دارد. در یک لایه، چندین نوع استخراج گر ویژگی<sup>51</sup> (لایه‌هایی که مقادیر ورودی را دریافت کرده، و به نوعی داده برای محاسبات تبدیل می‌کنند) وجود دارند. در شبکه‌ای که خود در حال یادگیری است و باید از گزینه‌های مختلفی برای حل وظایف استفاده کند، این نوع لایه‌بندی، به طور غیر مستقیم به عملکرد بهتر شبکه کمک می‌کند. این ماژول می‌تواند از ورودی‌ها مستقیماً در محاسبات خود استفاده کند، یا آن‌ها را مستقیماً جمع‌بندی کند.

در شکل ذیل که مربوط به معماری گوگل نت می باشد یک بلوک ماژول آغازین با رنگ قرمز مشخص شده است و همچنین دو دسته بند کمکی نیز با رنگ سبز مشخص شده اند.

<sup>51</sup> Feature Extractor

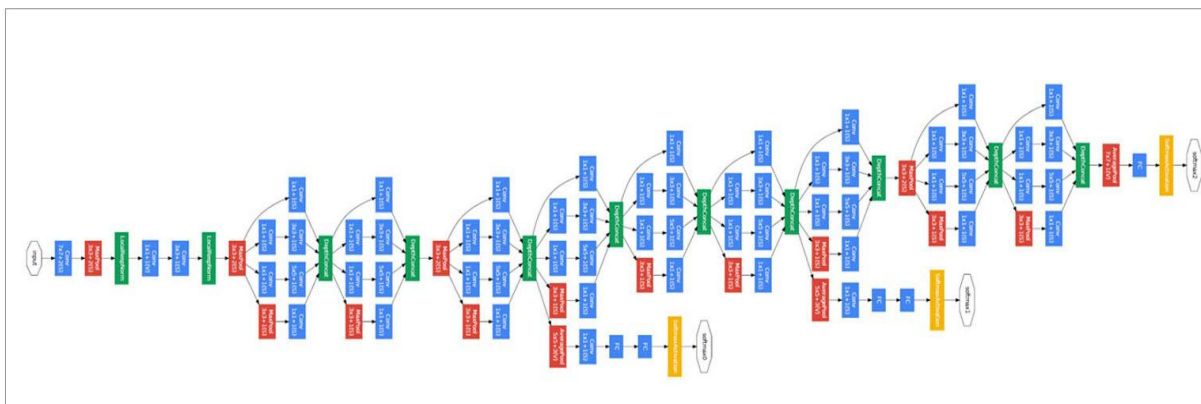




شکل 12- بلوک های مربوط به ماژول آغازین و دسته بند های کمکی

در شکل ذیل معماری کلی گوگل نت به همراه اطلاعات مربوط آن را می توان مشاهده کرد. همانطور که در تصویر ذیل مشخص می باشد در این معماری از هر دو نوآوری دسته بند های کمکی (مستطیل های زرد نشان دهنده دسته بندهای هستند که همانطور که در شکل مشخص می باشد این معماری دارای دو دسته بند کمکی می باشد که کمک می کند با وجود تعداد لایه های زیاد این معماری مشکل محو شدگی گرادیان به وجود نیاید) و همچنین ماژول آغازین (همانطور که در شکل بالا پیداست از چندین ماژول آغازین در این معماری استفاده شده است که به یادگیری بهتر و قابلیت تعمیم دهی<sup>52</sup> بهتر این معماری کمک می کند) استفاده شده است.

<sup>52</sup> Generalization



شکل 13- شکل کلی معماری گوگل نت

از مزایای گوگل نت می‌توان موارد زیر را نام برد:

- یادگیری گوگل نت بسیار سریع است.
- یادگیری گوگل نت بهتر از معماری های قبلی می باشد.
- گوگل نت دچار محو دگی گرادیان نمی شود.
- حجم یک مدل گوگل نت از پیش تعلیم دیده، بسیار سبک است. حجم گوگل نت تنها 96 مگابایت است.

در جدول ذیل می توان دستاورد ها، پیکربندی، سال انتشار و مقام های پنج معماری معرفی شده در مسابقه تشخیص تصویری مجموعه داده ایمیج نت را مشاهده کرد.

دستاورد	پیکربندی	مقام	سال	روش
اولین شبکه ی کانولوشنی که بسیار مطرح شد	2 لایه کانولوشن + 3 لایه تماما متصل	-	1998	<b>LeNet</b>
معماری مهمی که باعث علاقه بسیاری از محققان در زمینه بینایی کامپیوتر شد	5 لایه کانولوشن + 3 لایه تماما متصل	اول	2012	<b>AlexNet</b>
معرفی قسمت تجسم در مدل ها و تجسم عملکرد هر یک از ویژگی ها در لایه های مختلف	5 لایه کانولوشن + 3 لایه تماما متصل	اول	2013	<b>ZFNet</b>
ارزیابی کاملی از شبکه با عمق افزایشی	13-15 لایه کانولوشن + 3 لایه تماما متصل	دوم	2014	<b>VGG</b>
افزایش عمق و عرض شبکه را بدون افزایش نیازمندی های محاسباتی	21 لایه کانولوشن + 1 لایه تماما متصل	اول	2014	<b>GoogLeNet</b>

جدول 1- خلاصه اطلاعات پنج معماری معرفی شده

## نتیجه گیری و مقایسه:

### بهترین دقت روی داده های تست:

در پروژه برای ارزیابی مدل ها در اولین قدم بهترین دقت هر کدام از مدل ها روی داده های تست را باید ارزیابی کرد. (البته برای گوگل نت از دو بهینه ساز مختلف استفاده شده است که نتایج نشان داد که بهینه ساز<sup>53</sup> آدام<sup>54</sup> در این مورد عملکرد بهتری از بهینه ساز گرادینان نزولی تصادفی<sup>55</sup> داشته است برای همین در ادامه از همین بهینه ساز استفاده شده است.) در ادامه بهترین دقت هر کدام از مدل ها روی داده تست را می توان مشاهده کرد:

- گوگل نت اس جی دی: 0.7900
- گوگل نت آدام: 0.80600
- لنت: 0.65400
- الکس نت: 0.75800
- زد اف نت: 0.65200
- وی جی جی: 0.67600

که همانطور که پیداست گوگل نت کاملاً برتر بوده و بهترین عملکرد را داشته است.

### نمودار تغییرات دقت:

یکی از معیار های مناسب برای سنجی معماری ها و مدل های مختلف ارزیابی نمودار تغییرات دقت آنها در طول آموزش آنها می باشد منظور از نمودار دقت هم نمودار دقت مربوط به داده های آموزش و دقت بر روی داده های آموزش می باشد و هم دقت بر روی داده های تست می باشد. (که البته در گوگل نت چون دو دسته بند کمی وجود دارد دقت آنها نیز در طول آموزش در نمودار نشان داده شده است) که از این طریق می توان فهمید بهترین دقت تست در کجا رخ داده و کجا دقت های تست و آموزش شروع به فاصله گرفتن از هم کرده اند (بیش برآزش<sup>56</sup>) و به طور کلی در طول آموزش دقت های تست و آموزش چگونه تغییر کرده اند. در نمودار های ذیل نمودار تغییرات دقت هر پنج مدل را می توان مشاهده کرد. (البته برای گوگل نت چون دو دسته بند کمی هم داریم نمودار مربوط به دقت روی داده آموزش و تست از هم جدا شده اند.)

---

<sup>53</sup> Optimizer

<sup>54</sup> Adam

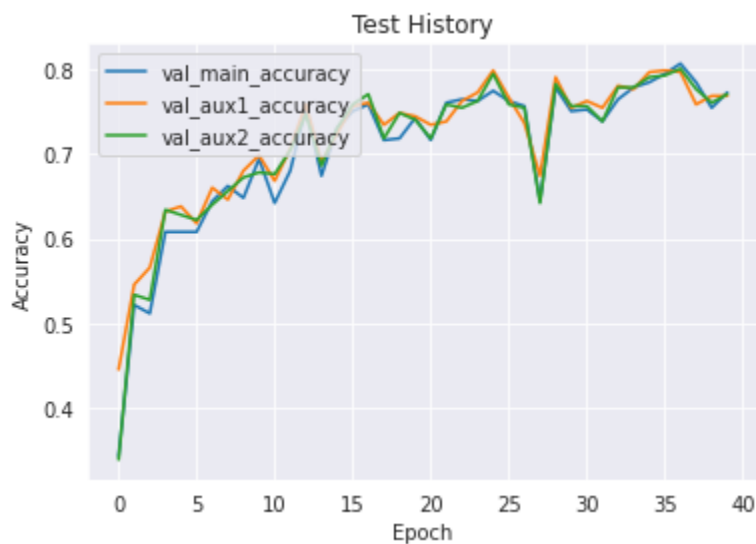
<sup>55</sup> Stochastic Gradient Descent

<sup>56</sup> Overfit

گول نت:

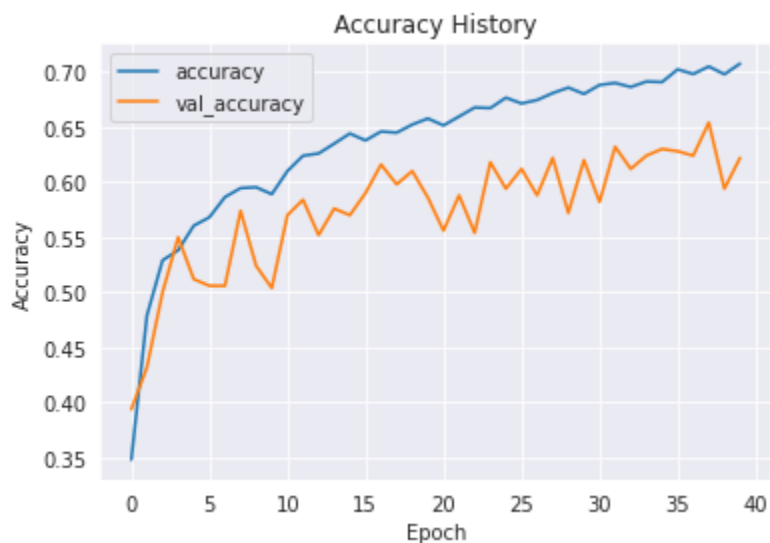


نمودار 2- تغییرات دقت معماری گول نت روی داده های آموزشی



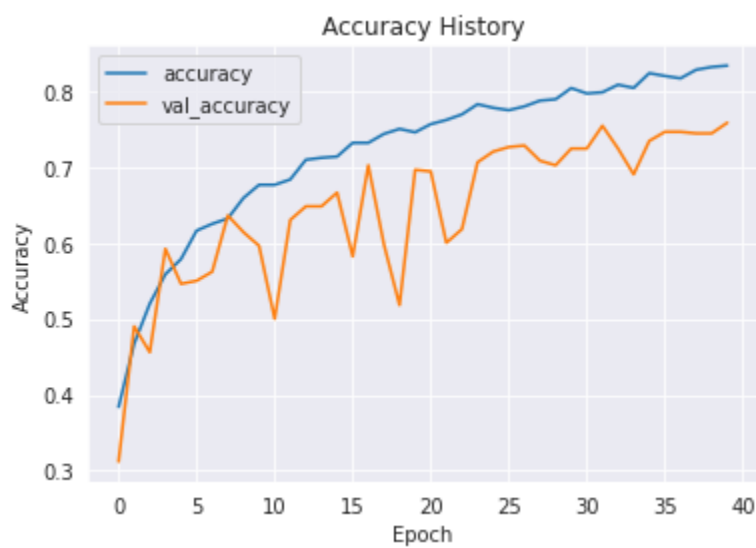
نمودار 3- تغییرات دقت معماری گول نت روی داده های تست

لنت:



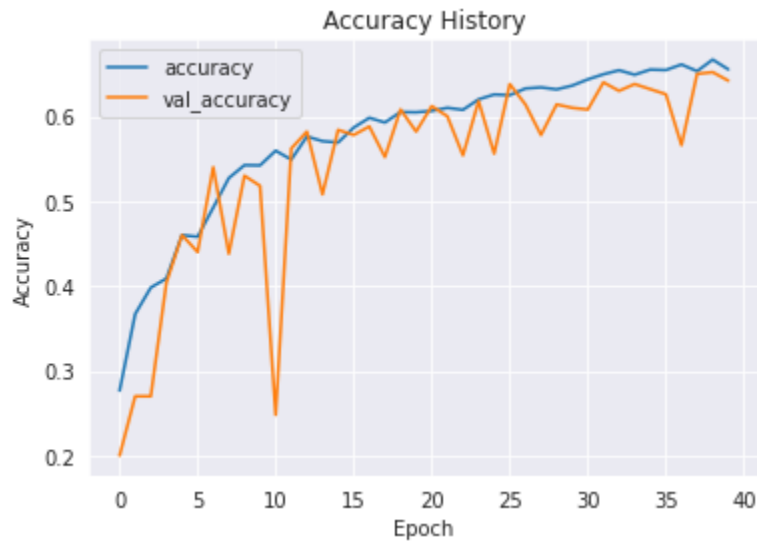
نمودار 4- تغییرات دقت معماری لنت روی داده های آموزشی و داده های تست

الکس نت:



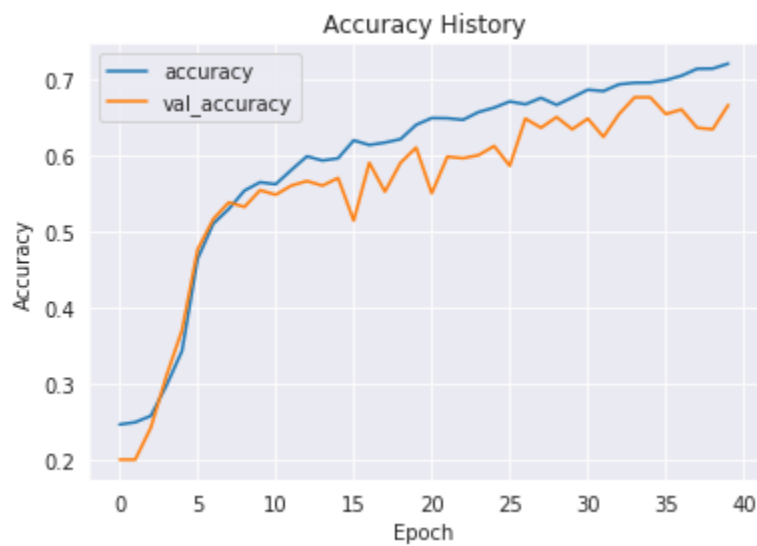
نمودار 5- تغییرات دقت معماری الکس نت روی داده های آموزشی و داده های تست

زد اف نت:



نمودار 6- تغییرات دقت معماری زد اف نت روی داده های آموزشی و داده های تست

وی جی جی 16:



نمودار 7- تغییرات دقت معماری وی جی جی 16 روی داده های آموزشی و داده های تست

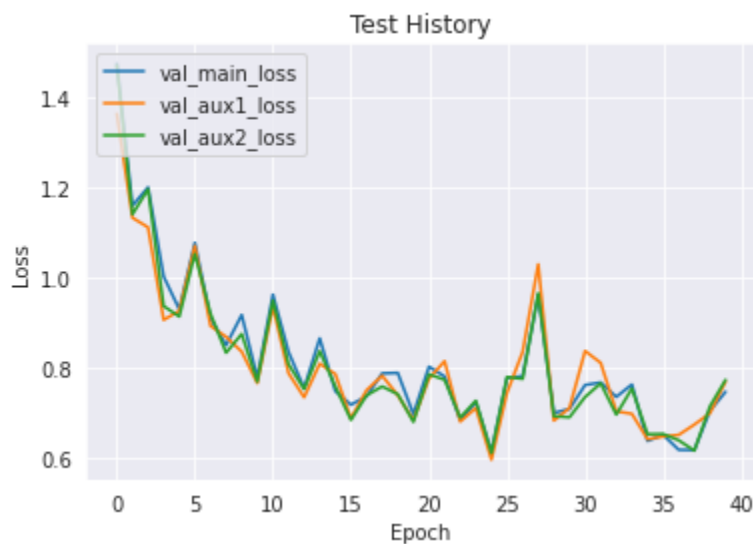
### نمودار تغییرات خروجی تابع ضرر:

نمودار مربوط به تغییرات خروجی تابع ضرر نیز مانند نمودار تغییرات دقت هستند و برای ارزیابی تغییرات خروجی تابع ضرر و اینکه آیا تابع ضرر و بهینه ساز استفاده شده مناسب می باشند یا خیر باید نمایش داده شوند که نمودار تغییرات تابع ضرر هر پنج معماری به شکل ذیل هستند. (باز هم برای گوگل نت چون دو دسته بند کمکی هم داریم نمودار مربوط به تغییرات خروجی تابع ضرر روی داده آموزش و تست از هم جدا شده اند).

گوگل نت:



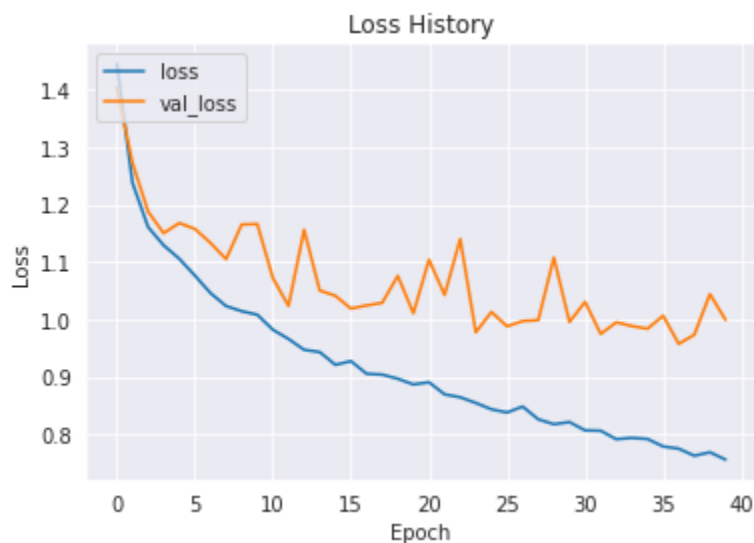
نمودار 8- تغییرات خروجی تابع ضرر معماری گوگل نت روی داده های آموزشی



نمودار 9- تغییرات خروجی تابع ضرر معماری گوگل نت روی داده های تست



لنت:



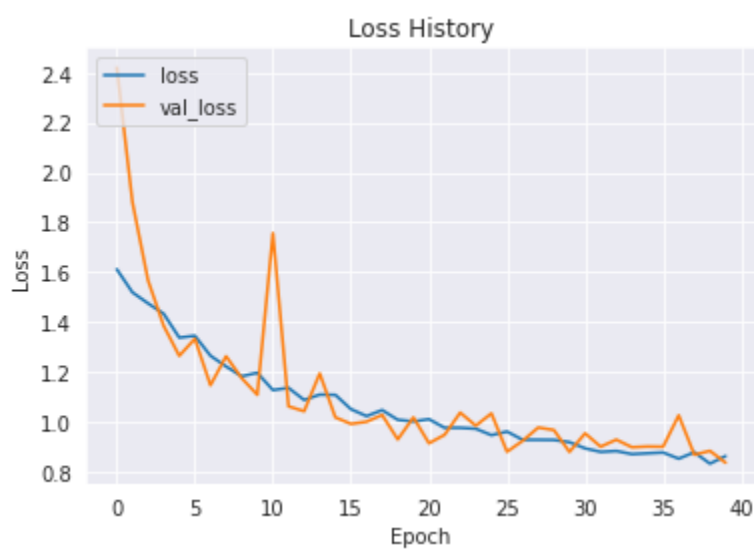
نمودار 10- تغییرات خروجی تابع ضرر معماری لنت روی داده های آموزشی و داده های تست

الکس نت:



نمودار 11- تغییرات خروجی تابع ضرر معماری الکس نت روی داده های آموزشی و داده های تست

زد اف نت:



نمودار 12- تغییرات خروجی تابع ضرر معماری زد اف نت روی داده های آموزشی و داده های تست

وی جی جی 16:



نمودار 13- تغییرات خروجی تابع ضرر معماری وی جی جی 16 روی داده های آموزشی و داده های تست

### ماتریس درهم ریختگی<sup>57</sup>:

همانطور که در کلاس درس دیدیم یکی از معیار های مهم و بسیار مناسب که برای سنجش مدل های مختلف می توان استفاده کرد ماتریس درهم ریختگی می باشد که نشان می دهد به اصلاح مدل ها هر کلاس را چند بار درست تشخیص داده اند و چند بار با کلاس های دیگر اشتباه گرفته اند که ما تواند اطلاعات زیادی در اختیار ما بگذارد تا بتوانیم مدل خود را اصلاح کنیم و بدانیم در کدام قسمت ها ضعف دارد و در کدام قسمت ها خوب عمل کرده است. در ادامه ماتریس درهم ریختگی مربوط به هر یک از مدل ها را می توان مشاهده کرد.

گوگل نت:

کلاس های پیش بینی شده توسط مدل						
برچسب داده ها		قاصدک	گل رز	گل لاله	گل آفتاب گردان	گل مروارید
	قاصدک	32	12	14	21	21
	گل رز	14	29	16	21	20
	لاله	14	28	33	14	11
	گل آفتاب گردان	20	18	16	26	20
	گل مروارید	13	13	25	25	24

جدول 2- ماتریس درهم ریختگی معماری گوگل نت

```
[[32 12 14 21 21]
 [14 29 16 21 20]
 [14 28 33 14 11]
 [20 18 16 26 20]
 [13 13 25 25 24]]
```

شکل 14- تصویر از نوت بوک برای ماتریس درهم ریختگی معماری گوگل نت

لنت:

کلاس های پیش بینی شده توسط مدل						
برچسب داده ها		قاصدک	گل رز	گل لاله	گل آفتاب گردان	گل مروارید
	قاصدک	16	33	18	15	18
	گل رز	8	31	16	18	27
	لاله	16	30	15	16	23
	گل آفتاب گردان	10	26	13	25	26
	گل مروارید	7	34	21	18	20

جدول 3- ماتریس درهم ریختگی معماری لنت

```
[[16 33 18 15 18]
 [ 8 31 16 18 27]
 [16 30 15 16 23]
 [10 26 13 25 26]
 [ 7 34 21 18 20]]
```

شکل 15- تصویر از نوت بوک برای ماتریس درهم ریختگی معماری لنت

الکس نت:

کلاس های پیش بینی شده توسط مدل						
برچسب داده ها		قاصدک	گل رز	گل لاله	گل آفتاب گردان	گل مروارید
	قاصدک	19	16	11	13	41
	گل رز	18	20	19	20	23
	لاله	20	19	20	19	22
	گل آفتاب گردان	19	22	15	16	28
	گل مروارید	22	24	16	19	19

جدول 4- ماتریس درهم ریختگی معماری الکس نت

```
[[19 16 11 13 41]
 [18 20 19 20 23]
 [20 19 20 19 22]
 [19 22 15 16 28]
 [22 24 16 19 19]]
```

شکل 16- تصویر از نوت بوک برای ماتریس درهم ریختگی معماری الکس نت

زد اف نت:

کلاس های پیش بینی شده توسط مدل						
برچسب داده ها		قاصدک	گل رز	گل لاله	گل آفتاب گردان	گل مروارید
	قاصدک	19	23	6	25	27
	گل رز	17	21	5	23	34
	لاله	22	20	1	19	38
	گل آفتاب گردان	15	24	11	24	26
	گل مروارید	19	18	8	19	36

جدول 5- ماتریس درهم ریختگی معماری زد اف نت

```
[[19 23 6 25 27]
 [17 21 5 23 34]
 [22 20 1 19 38]
 [15 24 11 24 26]
 [19 18 8 19 36]]
```

شکل 17- تصویر از نوت بوک برای ماتریس درهم ریختگی معماری زد اف نت

وی جی جی 16:

÷

کلاس های پیش بینی شده توسط مدل						
برچسب داده ها		قاصدک	گل رز	گل لاله	گل آفتاب گردان	گل مروارید
	قاصدک	17	32	18	20	13
	گل رز	15	22	17	24	22
	لاله	20	29	19	22	10
	گل آفتاب گردان	18	30	20	16	16
	گل مروارید	18	23	22	17	20

جدول 6- ماتریس درهم ریختگی معماری وی جی جی 16

```
[[17 32 18 20 13]
 [15 22 17 24 22]
 [20 29 19 22 10]
 [18 30 20 16 16]
 [18 23 22 17 20]]
```

شکل 18- تصویر از نوت بوک برای ماتریس درهم ریختگی معماری وی جی جی 16

### دقت سه کلاس برتر<sup>58</sup>:

یکی از معیار هایی که می توان به جای دقت در فقط تشخیص درست بودن هر کلاس استفاده کرد دقت سه کلاس برتر می باشد که به این صورت عمل می کند که اگر کلاس درست در بین سه کلاسی که مدل بالاترین احتمال را در خروجی به آنها داده است وجود داشته باشد آنگاه قابل قبول است و مانند این است که مدل کلاس را درست تشخیص داده ولی اگر غیر از این باشد قابل قبول نیست و مانند این است که مدل کلاس را اشتباه تشخیص داده. که دقت سه کلاس برتر هر یک از مدل ها به شرح ذیل می باشد:

- گوگل نت: 0.796
- لنت: 0.606
- الکس نت: 0.616
- زد اف نت: 0.622
- وی جی جی 16: 0.620

### جمع بندی:

در نهایت با توجه به مقایسه های انجام شده در بالا کاملاً مشخص می باشد که گوگل نت بهترین عملکرد را داشته و چه در دسته بندی های کلاس به کلاس چه دقت روی داده تست و چه در دقت سه کلاس برتر بهترین عملکرد را داشته و از بین موارد گفته شده بهترین انتخاب برای دسته بندی تصاویر می باشد.