

(написать заголовок)

(глава 1: актуальность статьи).(оценка по сравнению с предыдущими подходами).  
(достигнута цель или нет?).

Note that I refer to an updated version of the paper.(ссылка на нее). The updated version includes new works on the relevant topic and appendices. In this paper, authors used CNN to(or for) feature extraction and FC neural network to(or for) determine the visual similarity.(переписать первое предложение). In the last years, CNN to extract visual information have been proposed. However, visual similarity is commonly measured by computing a standard metric between vectors that are the output of the CNN. The main aim of the article is to use a neural network to define visual similarities between images and other researchers have not looked at, to the best of my knowledge, so I think that this article is relevant.

Naturally, method is far more expensive than computing the simple cosine similarity or Euclidean distance, therefore(?) we can expect that outperforms methods based on standard metric computations. ,otherwise.

(написать где-то в чем преимущества нейронной сети, над обычной метрикой). Naturally, feed-forwarding the two samples through the trained neural network is far more expensive than computing the cosine similarity or Euclidean distance, therefore(?) this approach must outperform methods based on standard metric computations. Otherwise there is no benefit to using a neural network. But the architecture of the CNN descriptor and preprocessing images also affect the result(?). I think that the goal of this work is performed, because the results obtained outperformed the results of other works with similar conditions.

(недостатки)

I think that(?) using metrics to set labels is not a good idea. Perhaps it is better than the naive classification, (результаты лучше чем больше отступ и это наводит на мысль что даже наивная классификация могла бы показать лучшие результаты.). but if we want to take into account the information that is lost when using metrics, we should choose another way of marking dataset(?). I will suggest my way later(?).

Написать, что из-за этой аппроксимации появляются противные три стадии и не получается end-to-end сети.

Еще недостаток, что плохо данные подготовили.

1)насколько актуальна поставленная задача с учетом всей предыдущей литературы:

В данной работе используется CNN для feature extraction выход из которого представляет вектор размерности .... и FC сеть для visual Similarity, выход из которого - это число, характеризующая степень близости двух картинок. CNN decoder для feature extraction использовалась и ранее, но в предыдущих работах использовали стандартные метрики для сравнения степени близости векторов, полученных после CNN decoder. Основная цель данной работы добиться лучших результатов за счет использования полносвязной нейронной сети для сравнения степени близости двух векторов, в предыдущих работах данный подход не рассматривался.

2)насколько логично, разумно, закончено и совершенно предлагаемое решение; Идея

мне кажется разумной, потому что стандартные метрики могут накладывать ограничения. (что-то еще надо написать тут). Но результаты не сильно превосходят, полученные в статье(Gordo et al.(2017)).

3 стадии обучения немного смущают, хотелось бы end-to-end.

3)в какой мере предложенное решение и полученные результаты соответствуют заявленным целям, в какой мере теоретический анализ и проведенные эксперименты подтверждают основные выводы работы?

4)в чем существенные фактические недочеты рассуждений, теоретических выкладок, предложенного решения и методологии экспериментов, если они имеются;

Заметки:

выбор классов изображений, то есть они считают, что классы похожи, если это одна и та же достопримечательность( а если на одном фото только колонна, а на втором все здание)

ошибка обобщения тоже играет роль( приложение а)

В релиз версии мы можем использовать FC нейронную сеть только для сравнения с теми фотографиями, где стандартная матрица дала высокие результаты.(?) ( Надо поисследовать границу порога)

Я бы изучил более подробно зависимость результатов стандартных метрик и результаты FC нейронной сети. Возможно, есть какой-то порог(слишком большая дистанция, например, для евклидовой метрикой), когда уже точно изображения непохожи и их не имеет смысла изучать с использованием нейронной сети. Это бы существенно могло бы увеличить производительность, так как большинство картинок из области поиска естественно не похожи на query image.

как понять по тестовым датасетам понимает ли он похожесть объектов или нет.

(не все методы поиска использовали). Standard image retrieval techniques that are commonly applied after cosine similarity computation, such as query expansion or image re-ranking, might also be applied on top of the similarity network.

подумать про то, что у них на выходе(почему нет сигмoиды).

## 1. Overview

In this paper, authors used CNN to feature extraction from image and FC neural network to determine the visual similarity. In the last years, CNN to extract visual information have been proposed. Furthermore, Gordo et al. (2017) used FC layers that replace the PCA projection. However, visual similarity is commonly measured by computing a standard metric between vectors that are the output of the CNN. The main aim of the article is to use a neural network to define visual similarities between images and other researchers have not looked at, to the best of my knowledge. Such an(!) approach will

show better results if the visual data presents any nonlinear interdependency. On the other hand, as already observed by the authors in appendix A, a subset of samples from the target dataset is required during training to learn a meaningful similarity function. So this article is quite interesting and relevant.

## 2. evaluation of results

Naturally, feed-forwarding the two samples through the trained neural network is far more expensive than computing the cosine similarity or Euclidean distance, hence this approach must outperform methods based on standard metric computations. Otherwise there is no benefit to using a neural network. But the architecture of the CNN descriptor (authors could use ResNet instead of a VGG16 network for image representation) and standard image retrieval techniques, such as query expansion or image re-ranking (authors did not use it), also have an impact on the result. Therefore, the goal of this paper is performed, because the results outperformed results of other works with similar conditions.

## 3. labeling datasets

I think that using cosine similarity to set labels is not a good idea. If we assume that the cosine similarity may give incorrect results, we should not use this metric to labeling dataset. Furthermore, this approach forces authors to use three stages of training. Perhaps even a naive classification would be better, because authors used Landmarks Gordo et al. (2017), where all images inside a class are visually related. Note that Gordo et al. (2017) obtained a set of pairwise scores between all image pairs and we can use this score for training procedure. Moreover, it can solve problem of approximating the K-nearest neighbors. We can construct a similar graph on a release set of images and compare query image with only a few nodes of each graph. As a result, we show a client all nodes of a graph, whose nodes turned out to be most similar to the original image. So I think that the naive classification or use scores for labeling dataset would be better than approximate cosine similarity with margin parameter.

## 4. Conclusion

In total, on the one hand authors presented a method that improved current results, but on the other hand