

# Компьютерное зрение

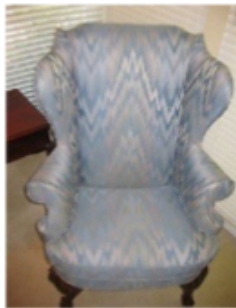
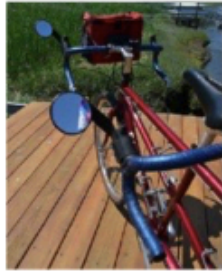
Практический курс  
Савельева Юлия Олеговна  
[i.o.saveleva.kpfu@gmail.com](mailto:i.o.saveleva.kpfu@gmail.com)  
2-й семестр, 07.05.2020 г.



# Image Retrieval

Stanford Online Products

## Query



## Retrieval



[https://cvgl.stanford.edu/projects/lifted\\_struct/](https://cvgl.stanford.edu/projects/lifted_struct/)

# Image Retrieval

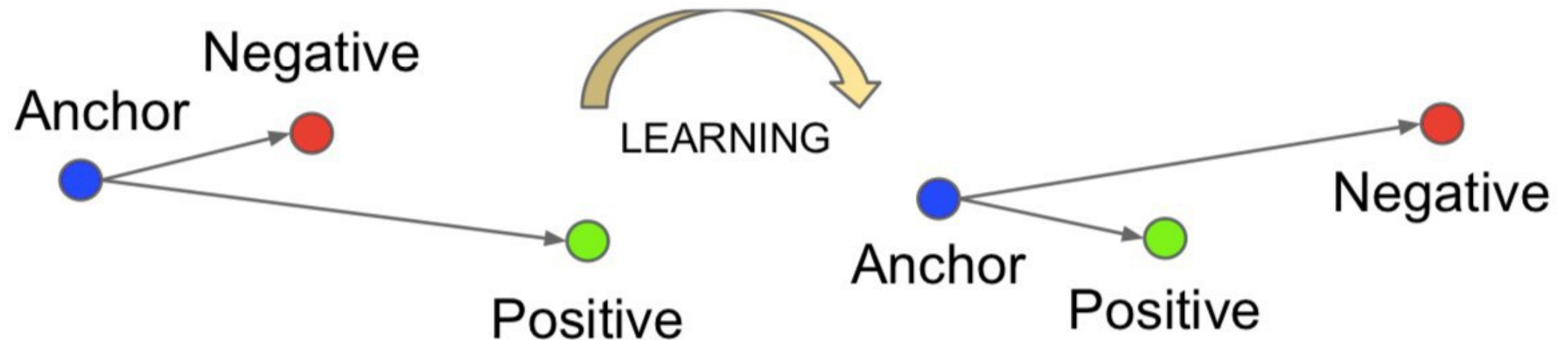
Архитектура



Необходимо использовать веса, обученные на задачу классификации.

# Image Retrieval

Triplet loss



$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 ,$$

$$\forall (f(x_i^a), f(x_i^p), f(x_i^n)) \in \mathcal{T} .$$

$$\text{Loss} = \sum_i^N \left[ \|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$



# Image Retrieval

## Triplet loss

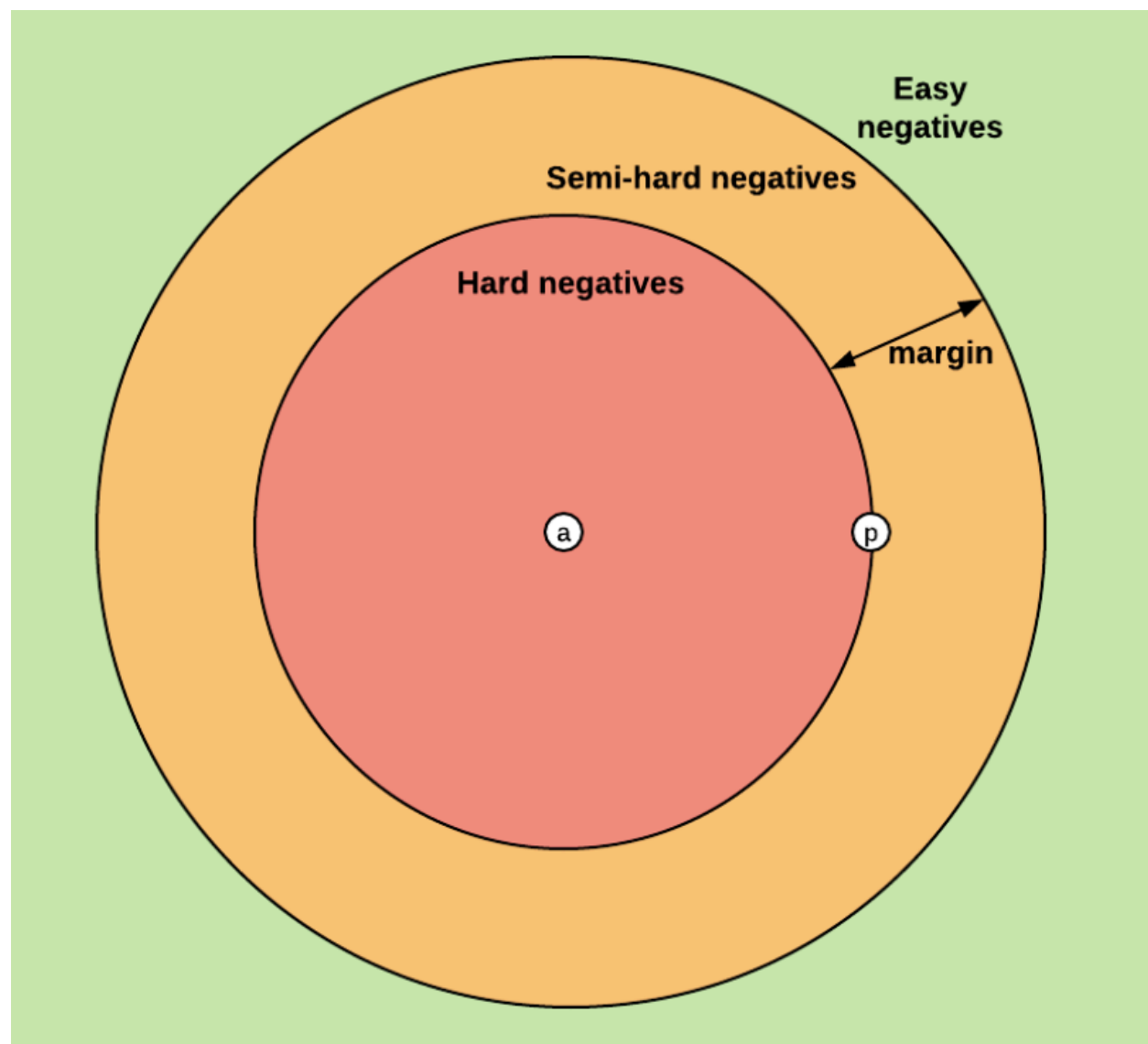
In order to ensure fast convergence it is crucial to select triplets that violate the triplet constraint in Eq. (1). This means that, given  $x_i^a$ , we want to select an  $x_i^p$  (*hard positive*) such that  $\operatorname{argmax}_{x_i^p} \|f(x_i^a) - f(x_i^p)\|_2^2$  and similarly  $x_i^n$  (*hard negative*) such that  $\operatorname{argmin}_{x_i^n} \|f(x_i^a) - f(x_i^n)\|_2^2$ .

Instead of picking the hardest positive, we use all anchor-positive pairs in a mini-batch while still selecting the hard negatives. We don't have a side-by-side comparison of hard anchor-positive pairs versus all anchor-positive pairs within a mini-batch, but we found in practice that the all anchor-positive method was more stable and converged slightly faster at the beginning of training.

# Image Retrieval

## Triplet loss

- **easy triplets:** triplets which have a loss of 0, because  $d(a,p)+margin < d(a,n)$
- **hard triplets:** triplets where the negative is closer to the anchor than the positive, i.e.  $d(a,n) < d(a,p)$
- **semi-hard triplets:** triplets where the negative is not closer to the anchor than the positive, but which still have positive loss:  $d(a,p) < d(a,n) < d(a,p)+margin$



# Image Retrieval

Triplet loss

Selecting the hardest negatives can in practice lead to bad local minima early on in training, specifically it can result in a collapsed model (*i.e.*  $f(x) = 0$ ). In order to mitigate this, it helps to select  $x_i^n$  such that

$$\|f(x_i^a) - f(x_i^p)\|_2^2 < \|f(x_i^a) - f(x_i^n)\|_2^2 \ .$$

Для батча размером 64 необходимо взять 4 категории по 4 продукта из категории и по 4 фотографии каждого продукта.

Формирование триплетов производится следующим образом: используются все возможные комбинации anchor и positive примеров, а в качестве negative выбираются semi-hard примеры, описанные выше

# Image Retrieval

## Подсказки

У предобученной на задачу классификации модели нужно обучать на задачу Image Retrieval все слои нейронной сети (последний fully-connected(fc) нужно заменить на fc по размеру embedding и обучить с нуля)

Во время дообучения необходимо перевести Batch Normalization в режим test и не обучать его

Вычисление Loss необходимо реализовать в матричном виде, для того, чтобы обучение шло быстро

Код валидации будет предоставлен чуть позже, пока оценить качество embeddings можно с помощью T-SNE



# Image Retrieval

Подсчет точности

Query



Retrieval



Query



Retrieval



# На следующие занятия

Домашние задания

1. Object Localization with Key Points



2. Image Retrieval



3. Visualize embeddings with T-SNE

