Adversarial Metric learning

조민기

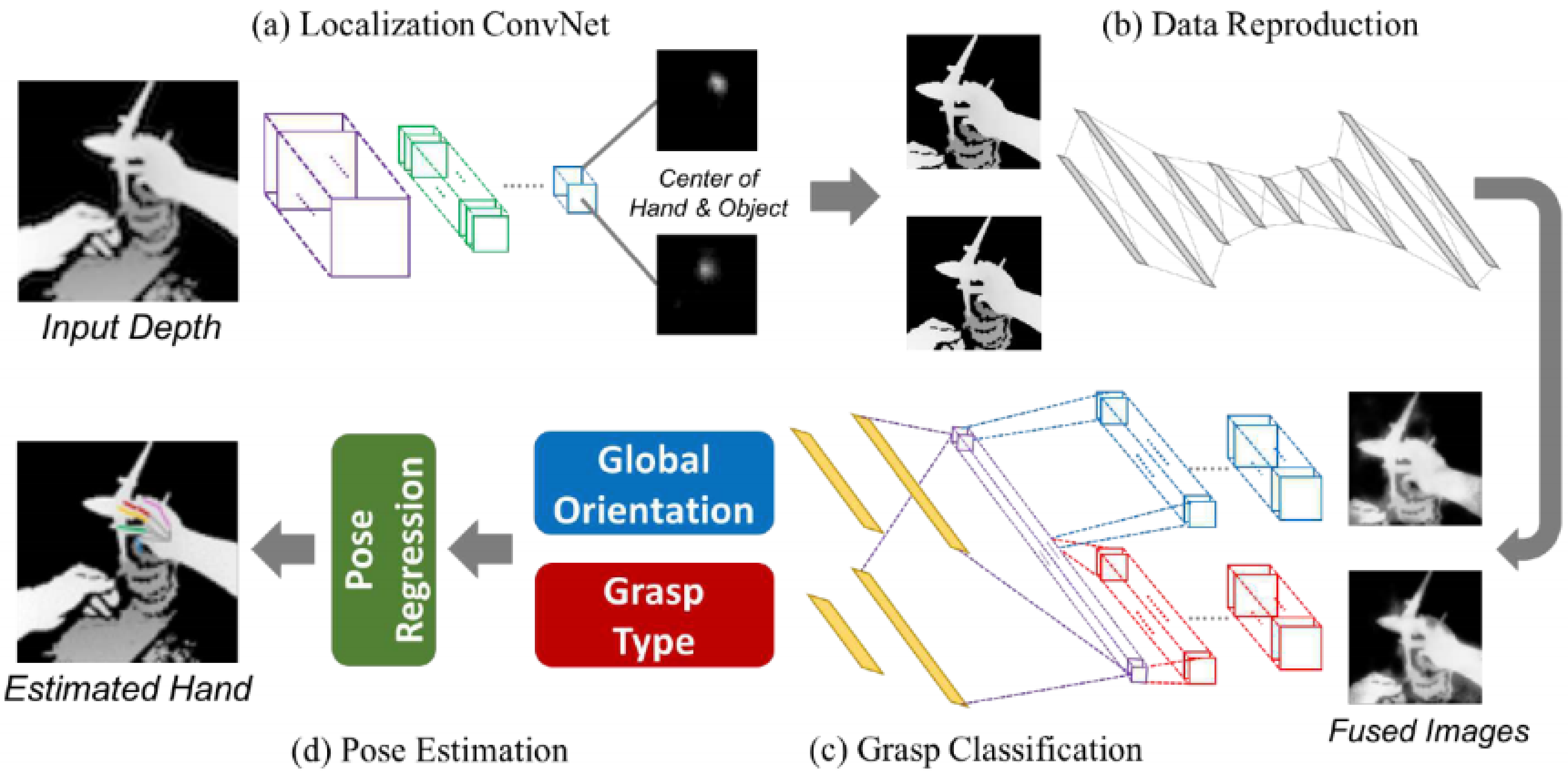
1

Introduction

PROBLEM

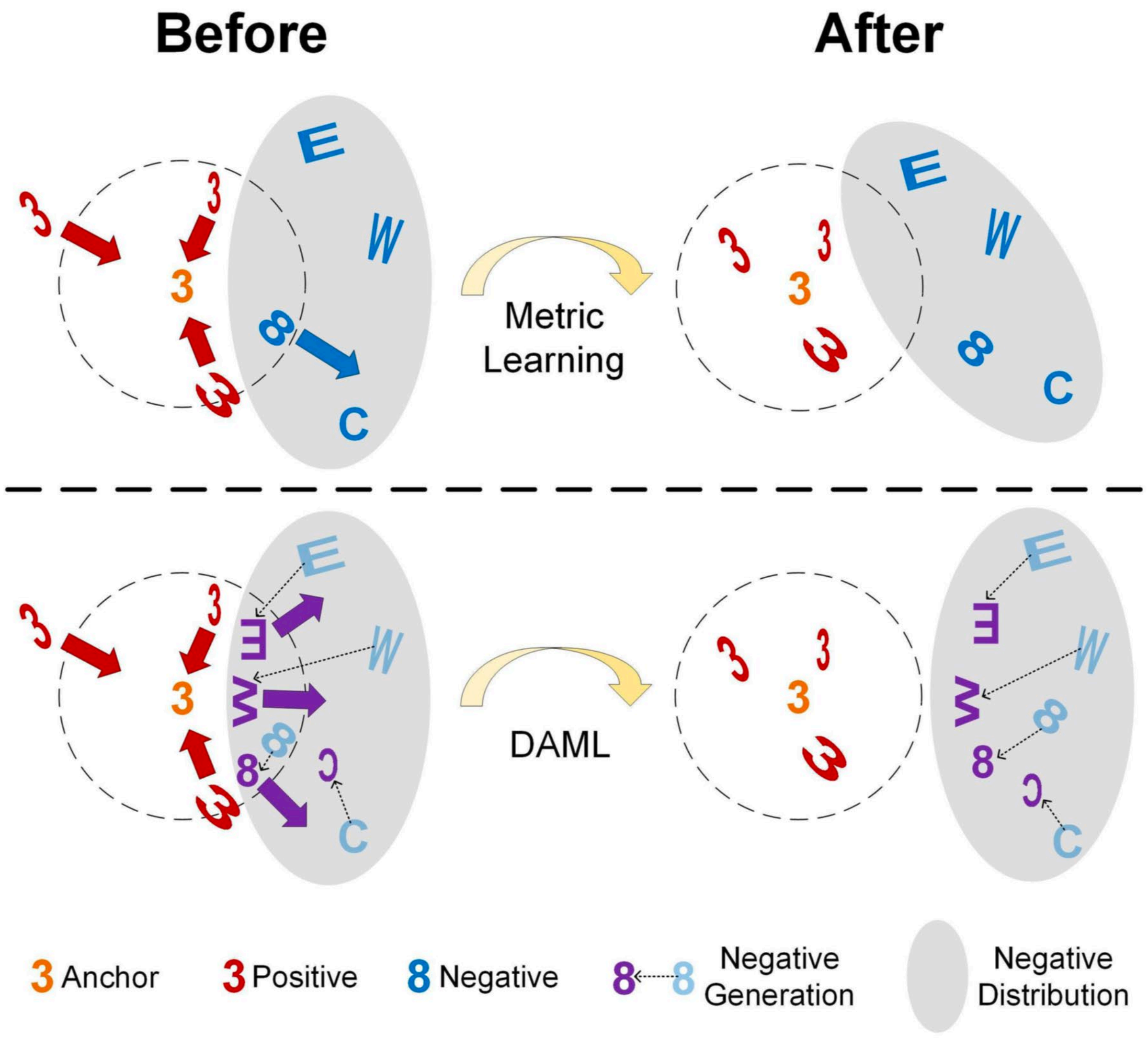
FUNCTION

- Object occlusion: multi-channel pipeline (hand / object)
- Dataset synthesis & Data reproduction



1 Introduction

- **Metric Learning**
- **Hard Negative Mining**
- **Adversarial Network**



2

Related work

Metric learning

Hard negative mining

Metric learning

- There are many matric to measure the similarity between two images such like PSNR, mPSNR, SSIM, MS-SSIM …
- However, these are just a mathematical measurement which is not intuitive.

이미지 유사도 측정 메트릭은 목적에 따라 많지
만 이들은 인간의 관점과 괴리가 크다.

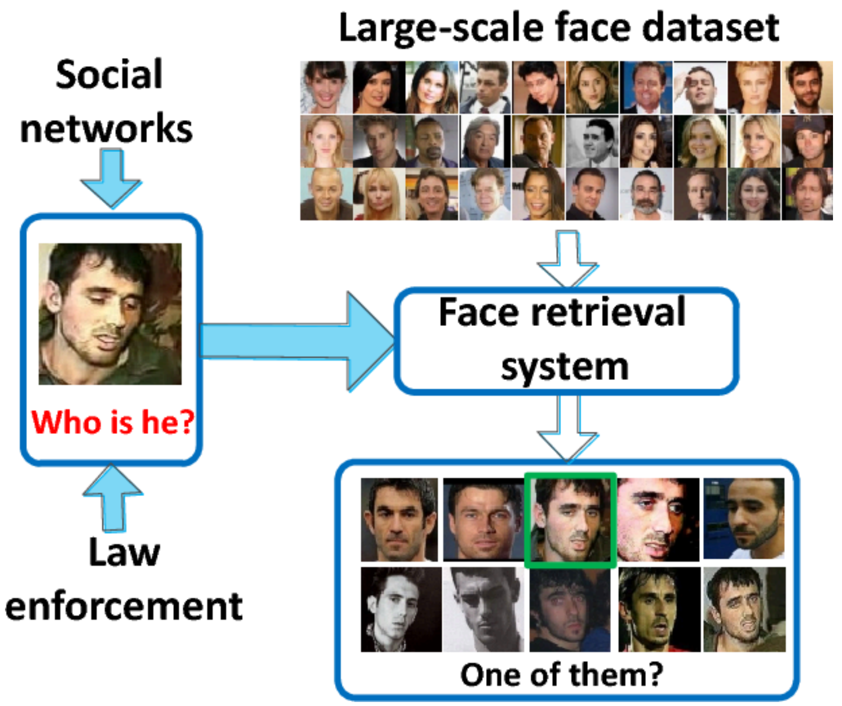
Metric learning

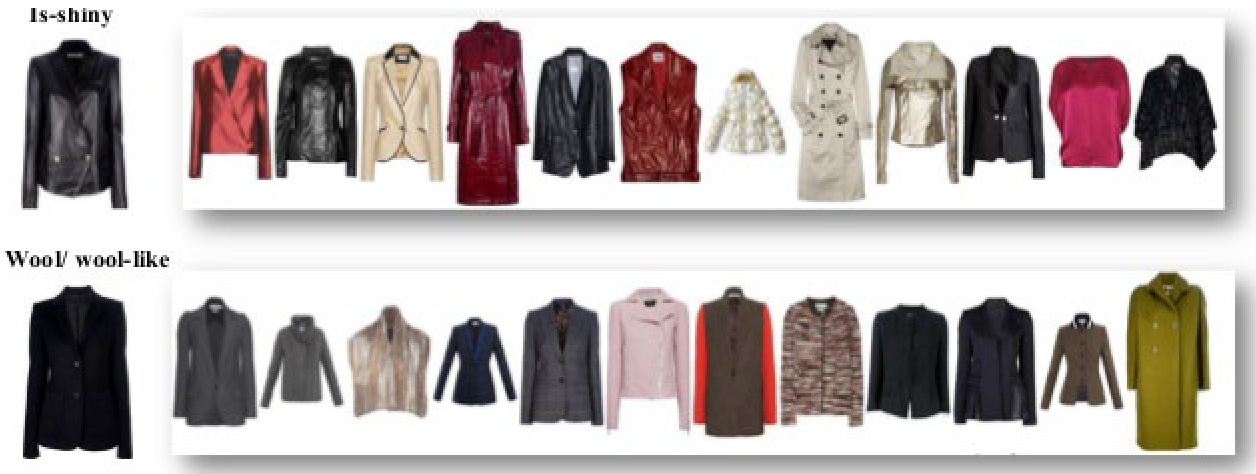
Despite more clear and realistic texture,
SRGAN image shows lower PSNR, SSIM value

SRResNet (23.53dB/0.7832) | SRGAN (21.15dB/0.6868) | original

Metric learning

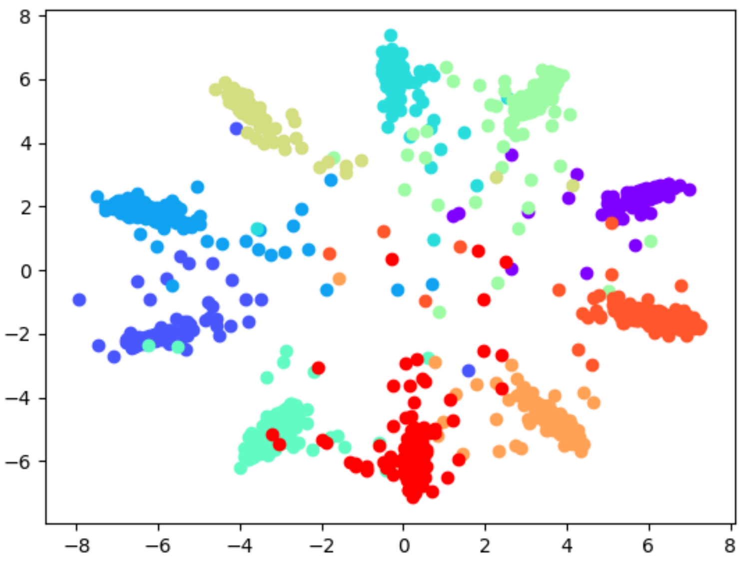
- The metric learning model learn the metric, such as the similarity of two images.
- Many face recognition service or fashion retrieval service are employing this model.





Metric learning - Demo

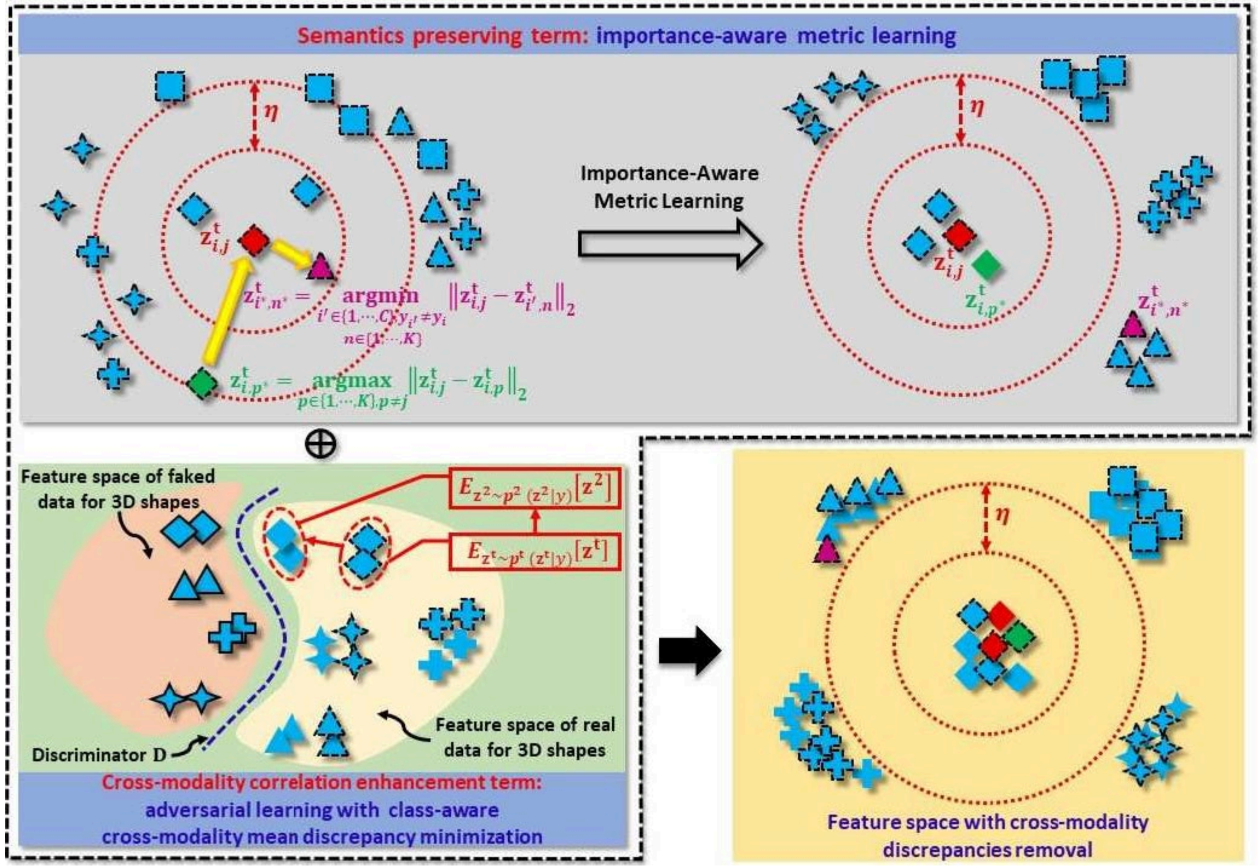
- The result of the base line code.
- The dimension reduction diagram of the embedding space (Fashion MNIST)
- Classification Accuracy 99.15%



Metric learning

- The basic objective is clustering the data in same category, and split the cluster which are in different category.

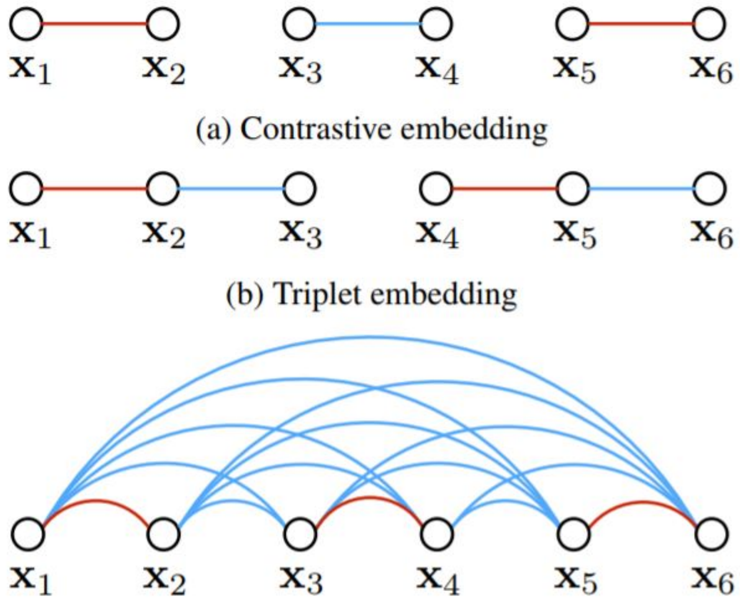
- Triplet loss is heavily used, and many works try to build a good embedding model based on their needs.



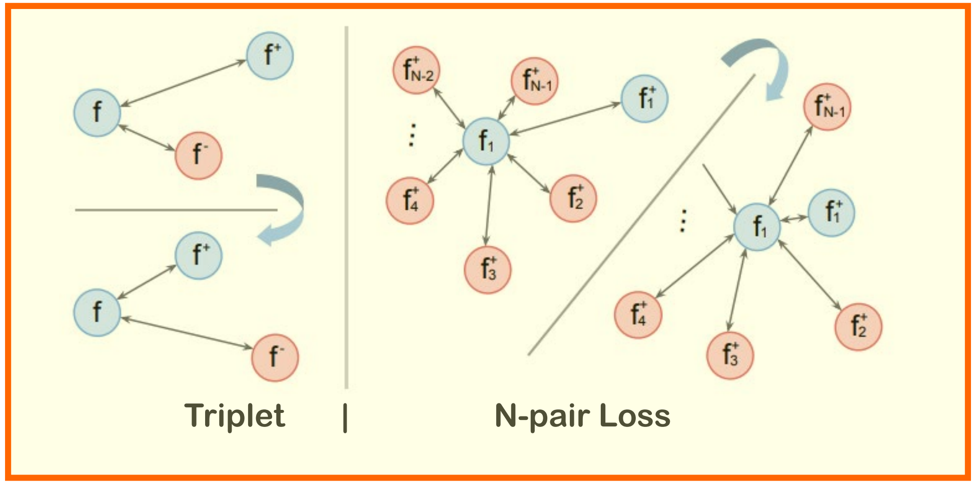
같은 카테고리의 데이터들과 가까워지고 다른 카테고리와 분리되도록 학습한다.

Metric Learning Loss

- Contrastive Loss
  Feed two examples
- Triplet Loss
  Less greedy than Contrastive loss
- Lifted structured loss
  Consider every point in the batch
- N-Pair Loss
  Consider N data from every category

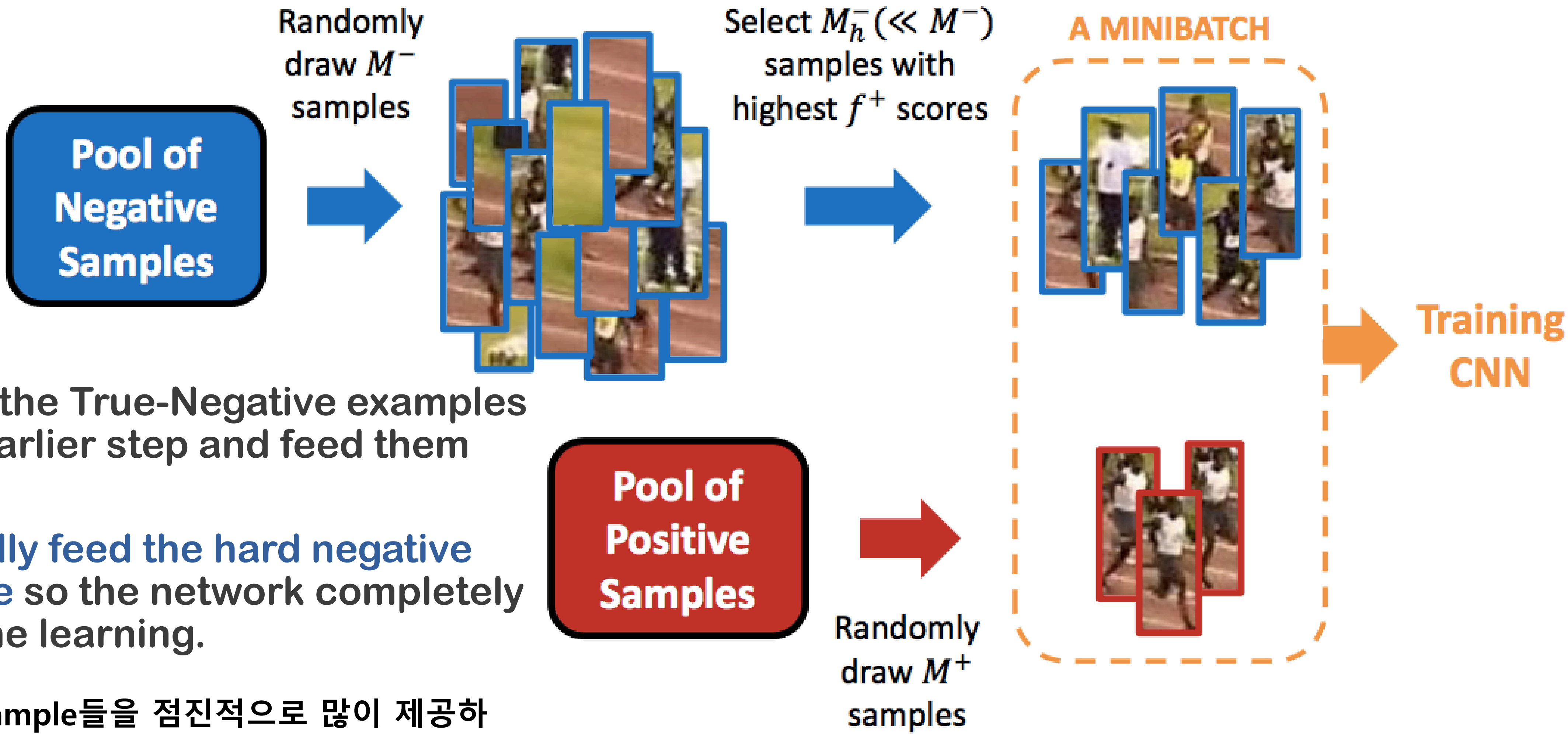


(a) Contrastive embedding
(b) Triplet embedding
(c) Lifted structured embedding



Improved Deep Metric Learning with Multi-class N-pair Loss Objective, NIPS 2016

Hard Negative mining



- Collect the True-Negative examples at the earlier step and feed them again.
- Gradually feed the hard negative example so the network completely finish the learning.

어려운 example들을 점진적으로 많이 제공하여 네트워크가 완전히 학습되도록 돕는다.

3

Model

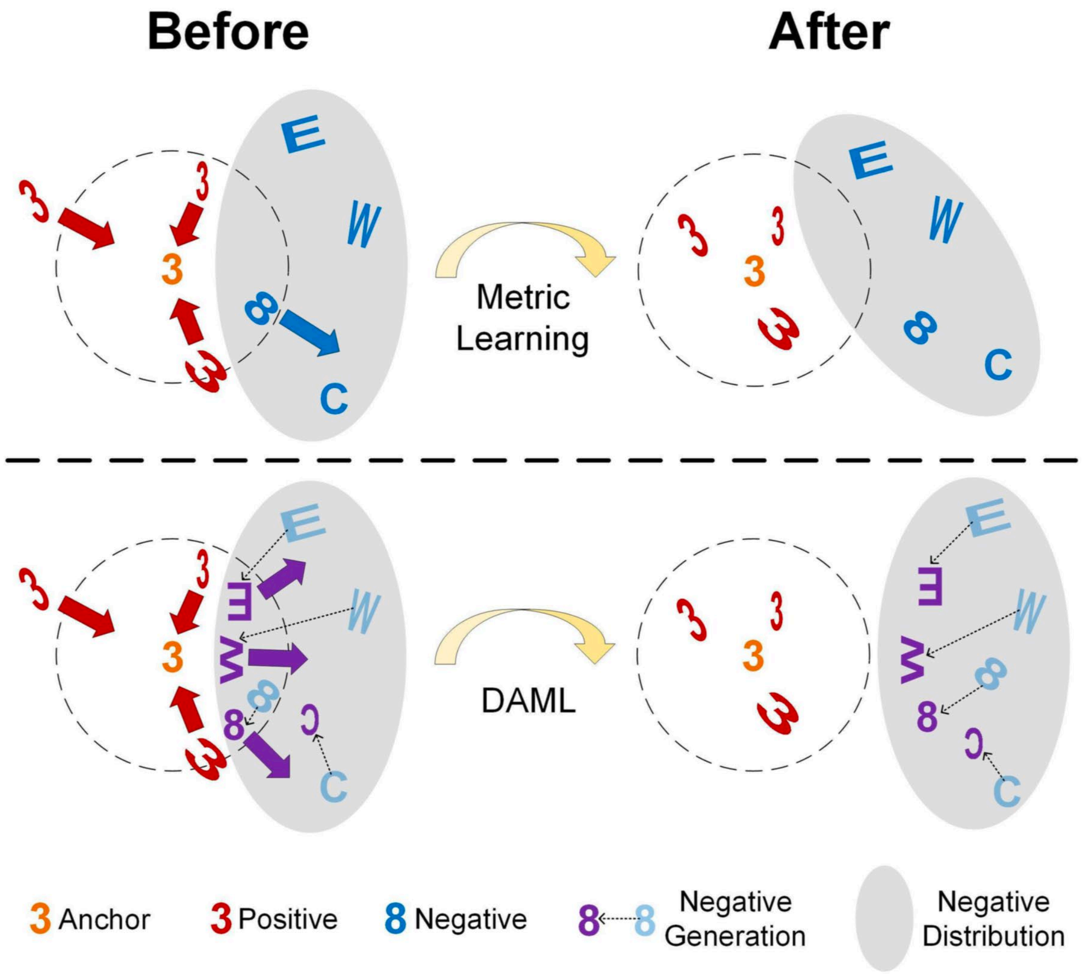
Model

Method

Main Idea

- Current works focused to embed the data well only with existing data.
- However, in this way, the Easy negative examples such like ‘W’, ‘E’ which might be ignored in earlier phase, could be a big threaten with some modification.
- Therefore, this paper Generate the Hard negative examples with existing easy negative samples in adversarial way.

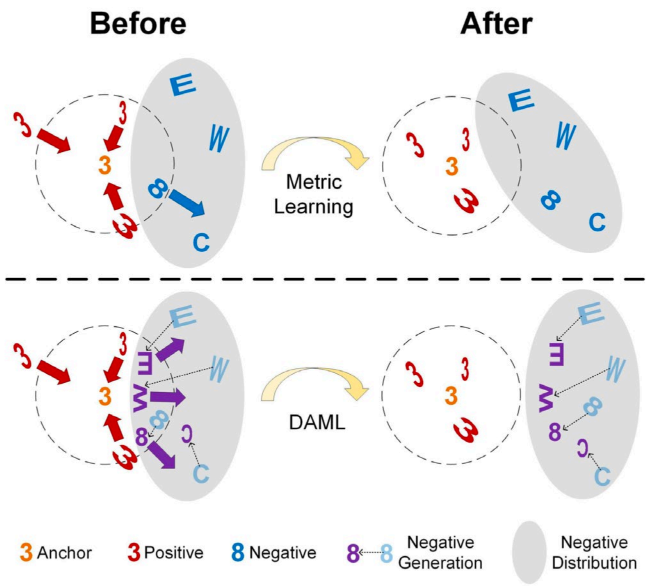
현재 기술은 존재하는 데이터만을 갖고 학습을 한다. 하지만 이 경우 W, E처럼 앞에서 Easy negativ였던 데이터가 조금의 변형으로 hard negative가 될 수 있음에도 이 경우를 무시하는 단점이 있다. 이를 GAN으로 방지한다.

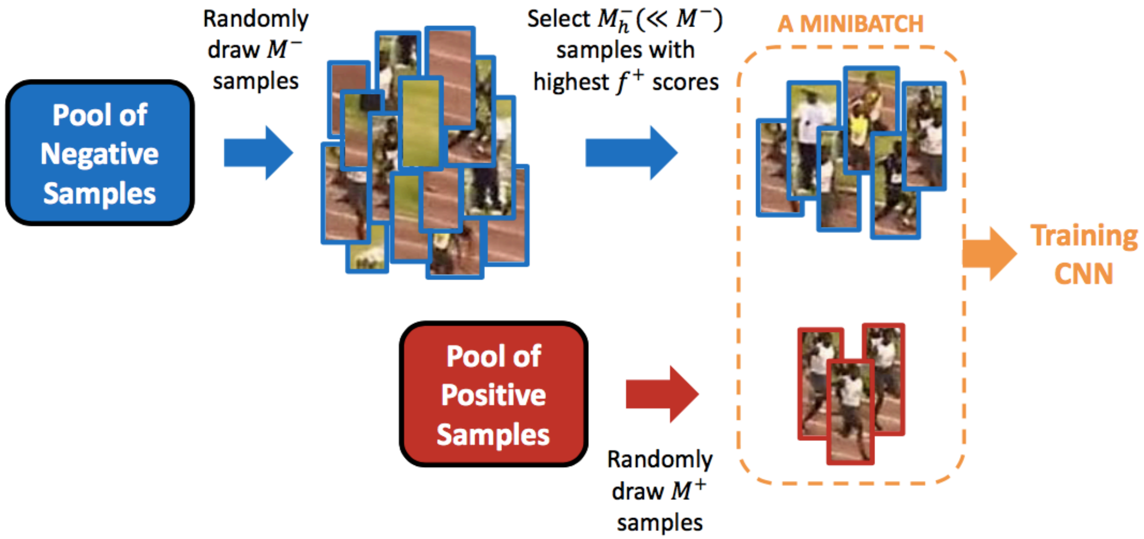


Contribution

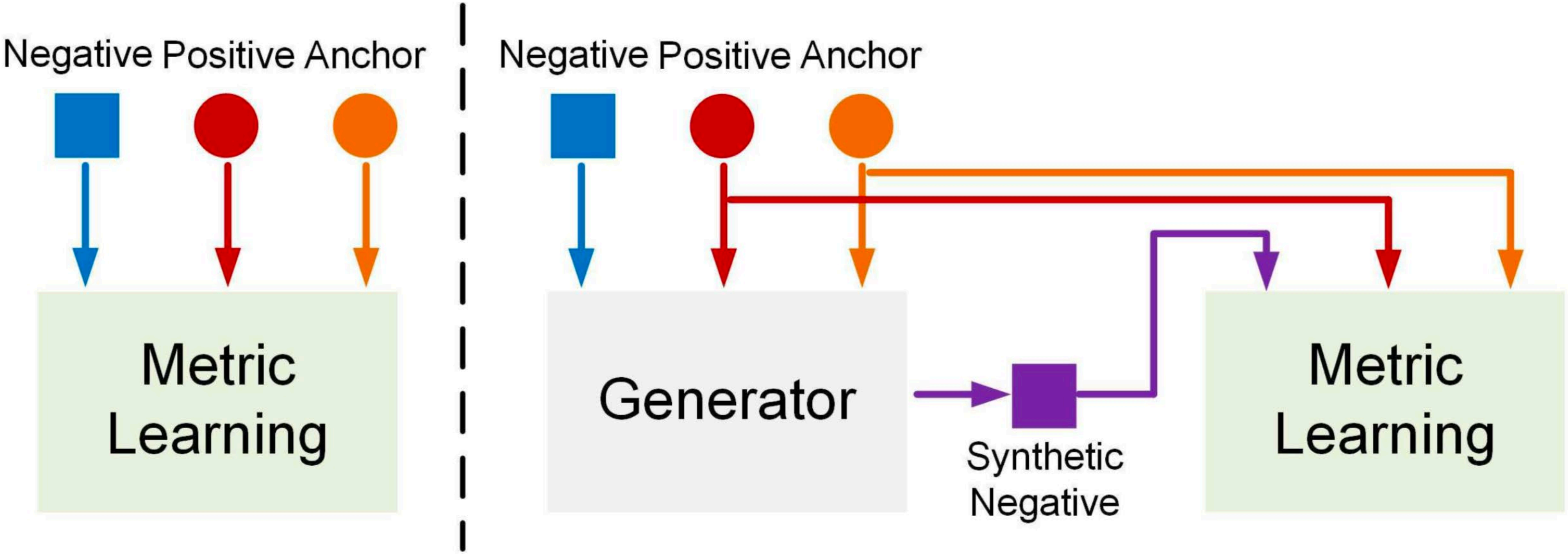
- Hard negative mining uses only with existing data. Therefore, the samples which was easy example could be ignored later.
- In contrast, DAML consider the potential threaten by generating the adversarial examples.
- Moreover, DAML does not conflict with hard negative mining because this method generates more negative example, not select the useful existing example.

쓸모있는 데이터를 다시 사용하는게 아닌, Negative 예시를 더 생성해낸다는 점에서 Hard negative mining과 다르다.



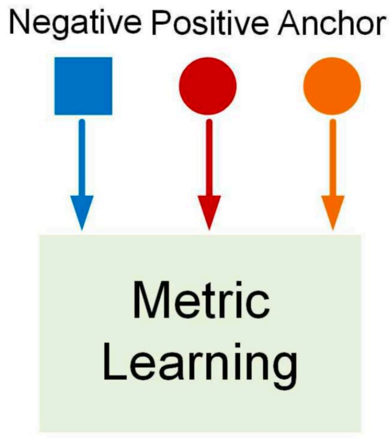


Proposed Method



Flow

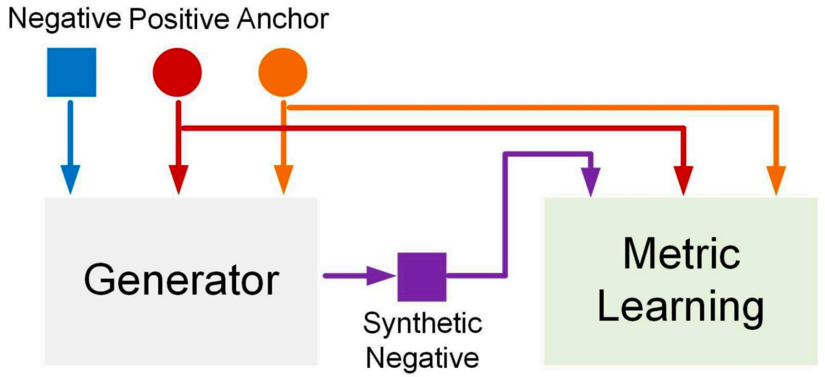
- **Supervised (current)Metric learning**



Embedding

$$\theta_f = \arg\min_{\theta_f} J_{\mathrm{m}}(\theta_f; \mathbf{x}_i, \mathbf{x}_i^+, \mathbf{x}_i^-, f)$$

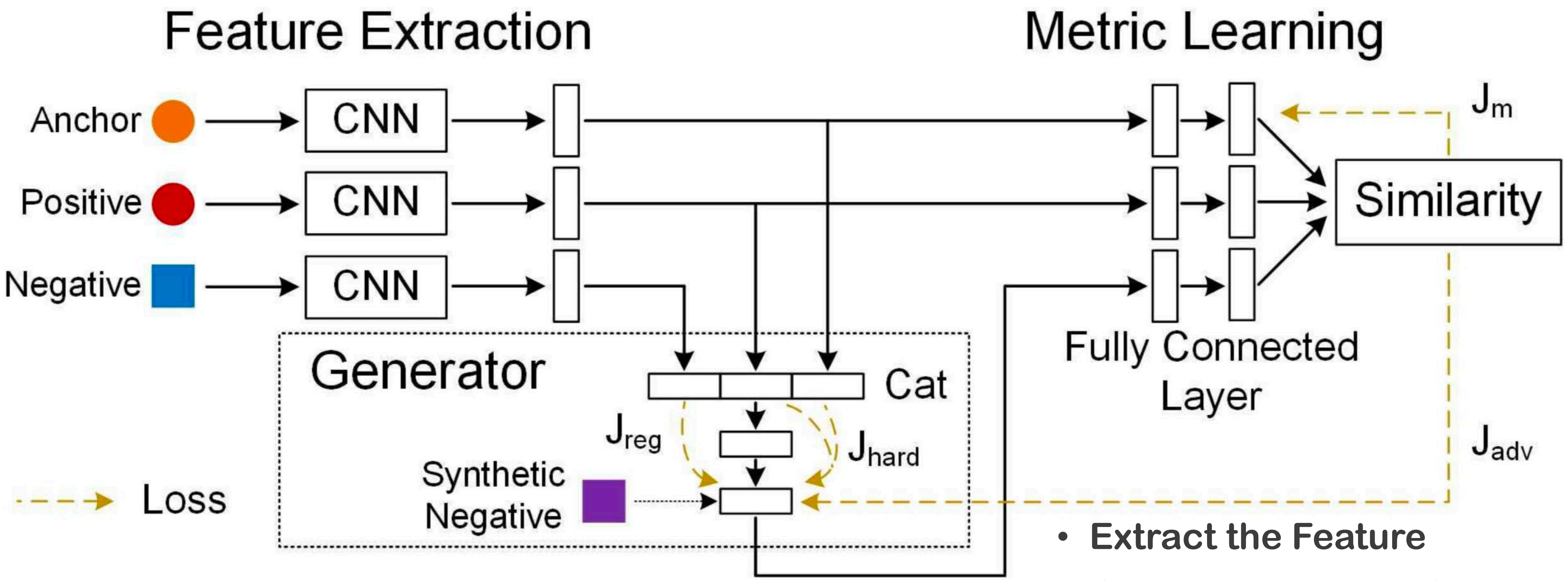
Weight



$$\theta_f^a = \arg\min_{\theta_f} J_{\mathrm{m}}(\theta_f; \mathbf{x}_i, \mathbf{x}_i^+, \widetilde{\mathbf{x}}_i^-, f)$$

쓸모있는 데이터를 다시 사용하는게 아닌, Negative 예시를 더 생성해낸다는 점에서 Hard negative mining과 다르다.

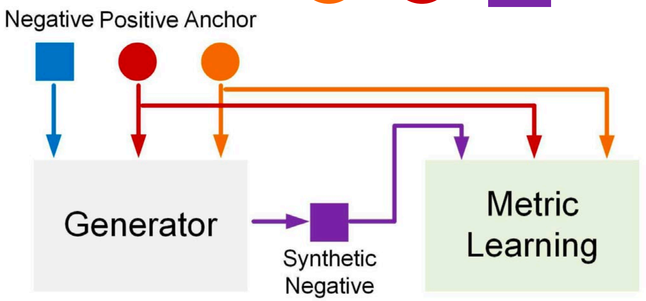
Flow



- Extract the Feature
- Generate Negative feature
- Embed through the FCN

Hard Negative Generator

$$\theta_f^a = \arg\min_{\theta_f} J_{\mathrm{m}}(\theta_f; \mathbf{x}_i, \mathbf{x}_i^+, \widetilde{\mathbf{x}}_i^-, f)$$



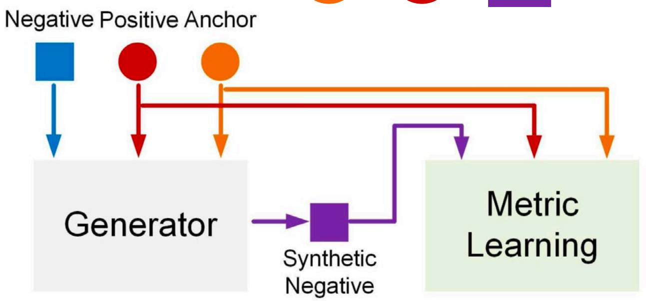
$$\widetilde{\mathbf{x}}_i^- = G(\theta_g; \mathbf{x}_i^-, \mathbf{x}_i, \mathbf{x}_i^+)$$

$$\begin{aligned}
\min_{\theta_g} J_{\mathrm{gen}} &= J_{\mathrm{hard}} + \lambda_1 J_{\mathrm{reg}} + \lambda_2 J_{\mathrm{adv}} \\
&= \sum_{i=1}^{N} (||\widetilde{\mathbf{x}}_i^- - \mathbf{x}_i||_2^2 + \lambda_1 ||\widetilde{\mathbf{x}}_i^- - \mathbf{x}_i^-||_2^2 \\
&+ \lambda_2 [D(\widetilde{\mathbf{x}}_i^-, \mathbf{x}_i)^2 - D(\mathbf{x}_i^+, \mathbf{x}_i)^2 - \alpha]_+)
\end{aligned}$$

3 Model

Hard Negative Generator

$$\theta_f^a = \arg\min_{\theta_f} J_{\mathrm{m}}(\theta_f; \mathbf{x}_i, \mathbf{x}_i^+, \widetilde{\mathbf{x}}_i^-, f)$$



$$\widetilde{\mathbf{x}}_i^- = G(\theta_g; \mathbf{x}_i^-, \mathbf{x}_i, \mathbf{x}_i^+)$$

$$\begin{aligned}\min_{\theta_g} J_{\mathrm{gen}} &= J_{\mathrm{hard}} + \boxed{\lambda_1 J_{\mathrm{reg}}} + \lambda_2 J_{\mathrm{adv}}\end{aligned}$$

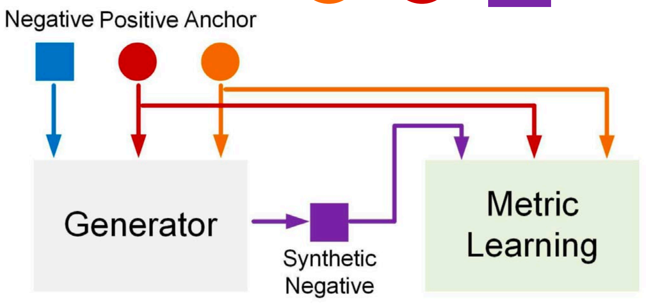
- Regularize the generator so the adversarial example does not go too far from the real negative example.

$$\begin{aligned} &= \sum_{i=1}^{N} (\|\widetilde{\mathbf{x}}_i^- - \mathbf{x}_i\|_2^2 + \boxed{\lambda_1 \|\widetilde{\mathbf{x}}_i^- - \mathbf{x}_i^-\|_2^2} \\ &+ \lambda_2 [D(\widetilde{\mathbf{x}}_i^-, \mathbf{x}_i)^2 - D(\mathbf{x}_i^+, \mathbf{x}_i)^2 - \alpha]_+)\end{aligned}$$

3 Model

Hard Negative Generator

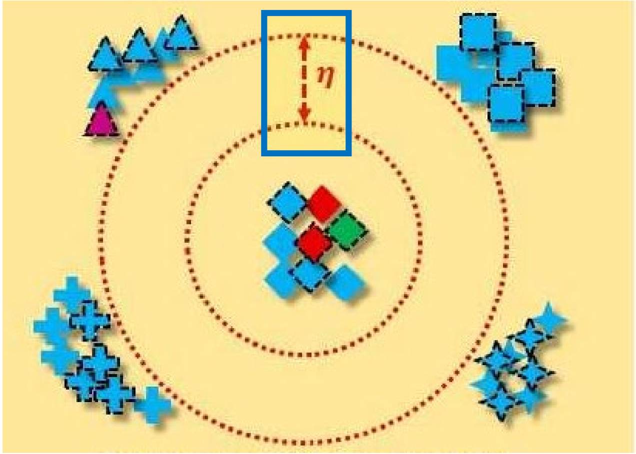
$$\theta_f^a = \arg\min_{\theta_f} J_{\mathrm{m}}(\theta_f; \mathbf{x}_i, \mathbf{x}_i^+, \widetilde{\mathbf{x}}_i^-, f)$$



$$\widetilde{\mathbf{x}}_i^- = G(\theta_g; \mathbf{x}_i^-, \mathbf{x}_i, \mathbf{x}_i^+)$$

$$\begin{aligned}\min_{\theta_g} J_{\text{gen}} &= J_{\text{hard}} + \lambda_1 J_{\text{reg}} + \lambda_2 J_{\text{adv}} \\ &= \sum_{i=1}^{N} \left(\|\widetilde{\mathbf{x}}_i^- - \mathbf{x}_i\|_2^2 + \lambda_1 \|\widetilde{\mathbf{x}}_i^- - \mathbf{x}_i^-\|_2^2 \right. \\ &+ \left. \lambda_2 [D(\widetilde{\mathbf{x}}_i^-, \mathbf{x}_i)^2 - D(\mathbf{x}_i^+, \mathbf{x}_i)^2 - \boxed{\alpha}]_+ \right)\end{aligned}$$

- The alpha margin of the cluster



4

Analyze

Experiment

Result

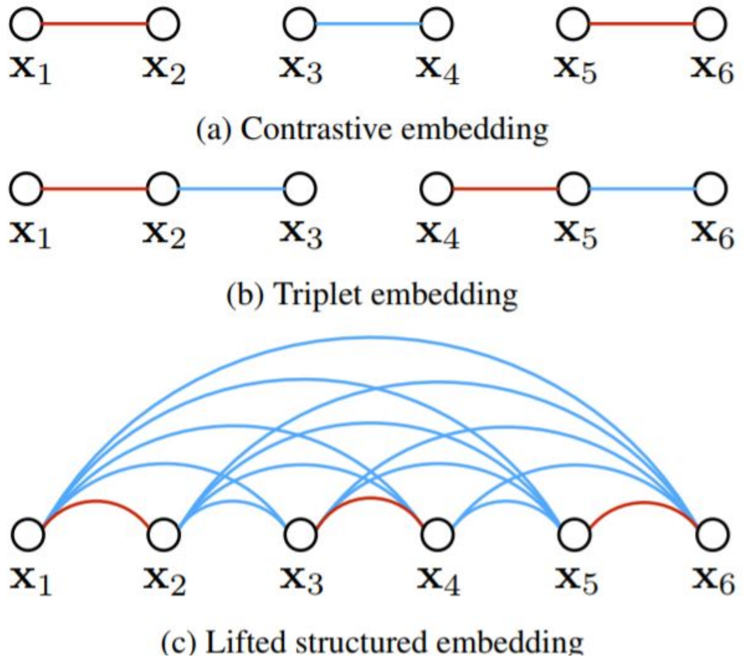
Experiment

- Contrastive Loss
  Feed two examples

- Triplet Loss
  Less greedy than Contrastive loss

- Lifted structured loss
  Consider every point in the batch

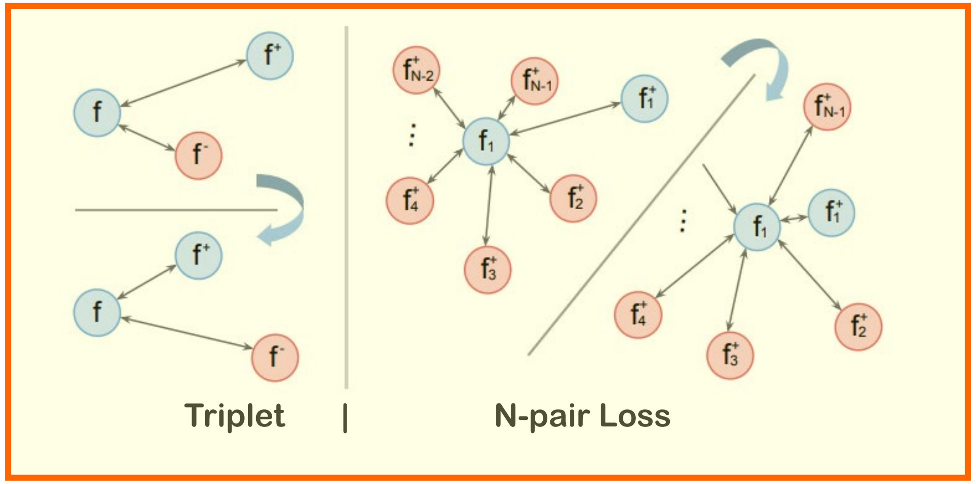
- N-Pair Loss
  Consider N data from every category



(a) Contrastive embedding

(b) Triplet embedding

(c) Lifted structured embedding



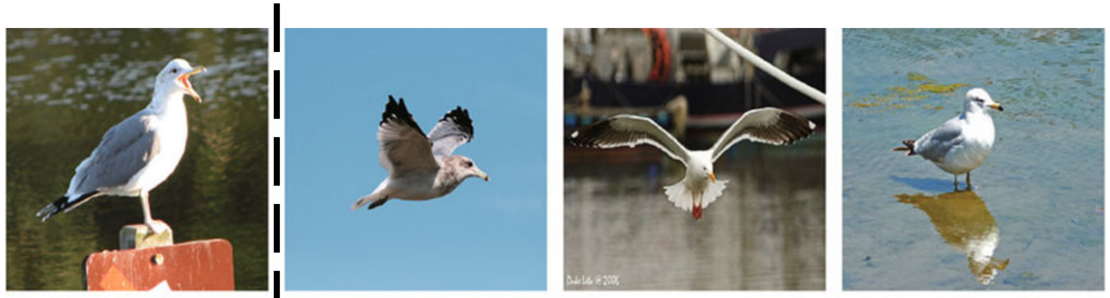
Improved Deep Metric Learning with Multi-class N-pair Loss Objective, NIPS 2016

4 Analize

Experiment

- With the Bird, Cars, Product image, conduct the image embedding on the metric space.
    - CUB200, Cars196, Stanford Online Product dataset

- Image retrieval with NN search



(a) Calilfornia gull

(b) Glaucous gull

Experiment

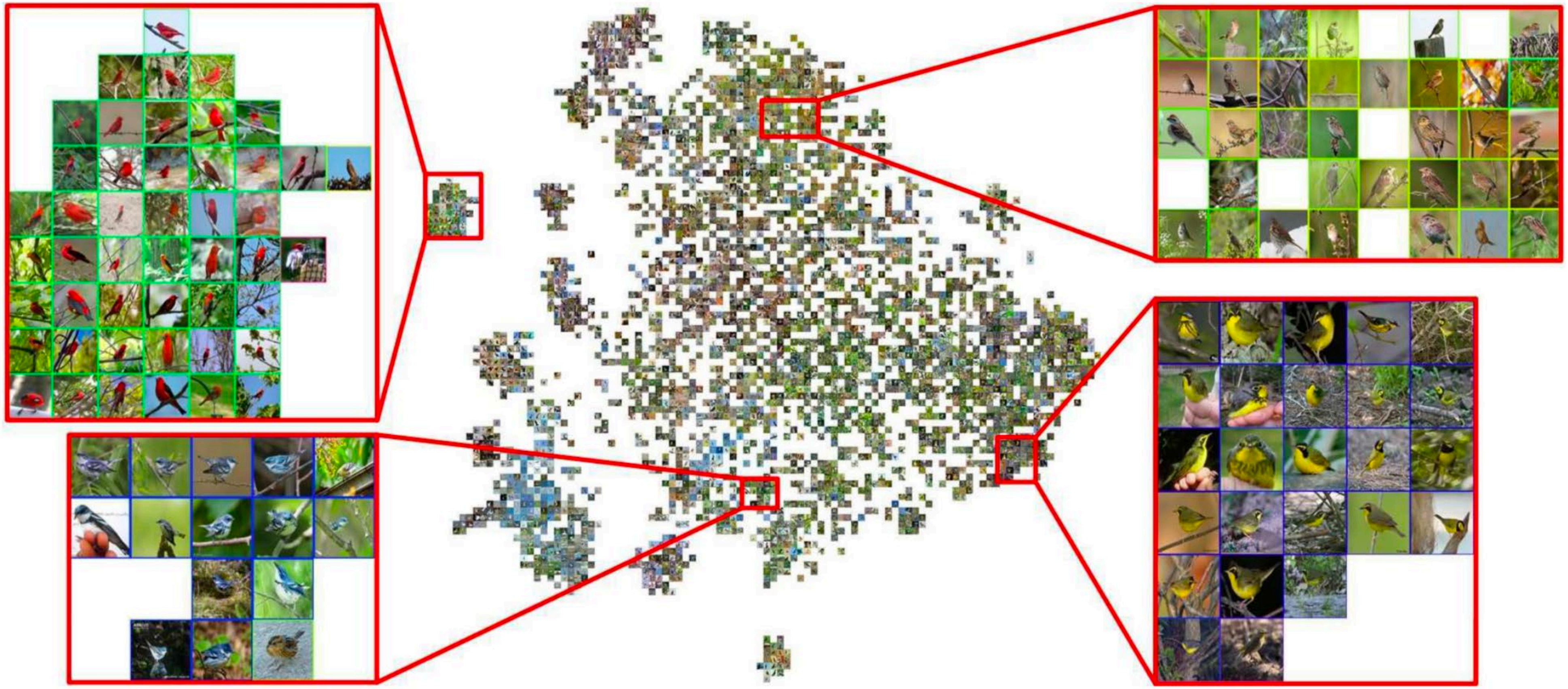


Figure 4. Visualization of the proposed DAML (N-pair) with Barnes-Hut t-SNE [37] on the CUB-200-2011 dataset, where the color of the border for each image represents the label. (Best viewed on a monitor when zoomed in.)

Experiment

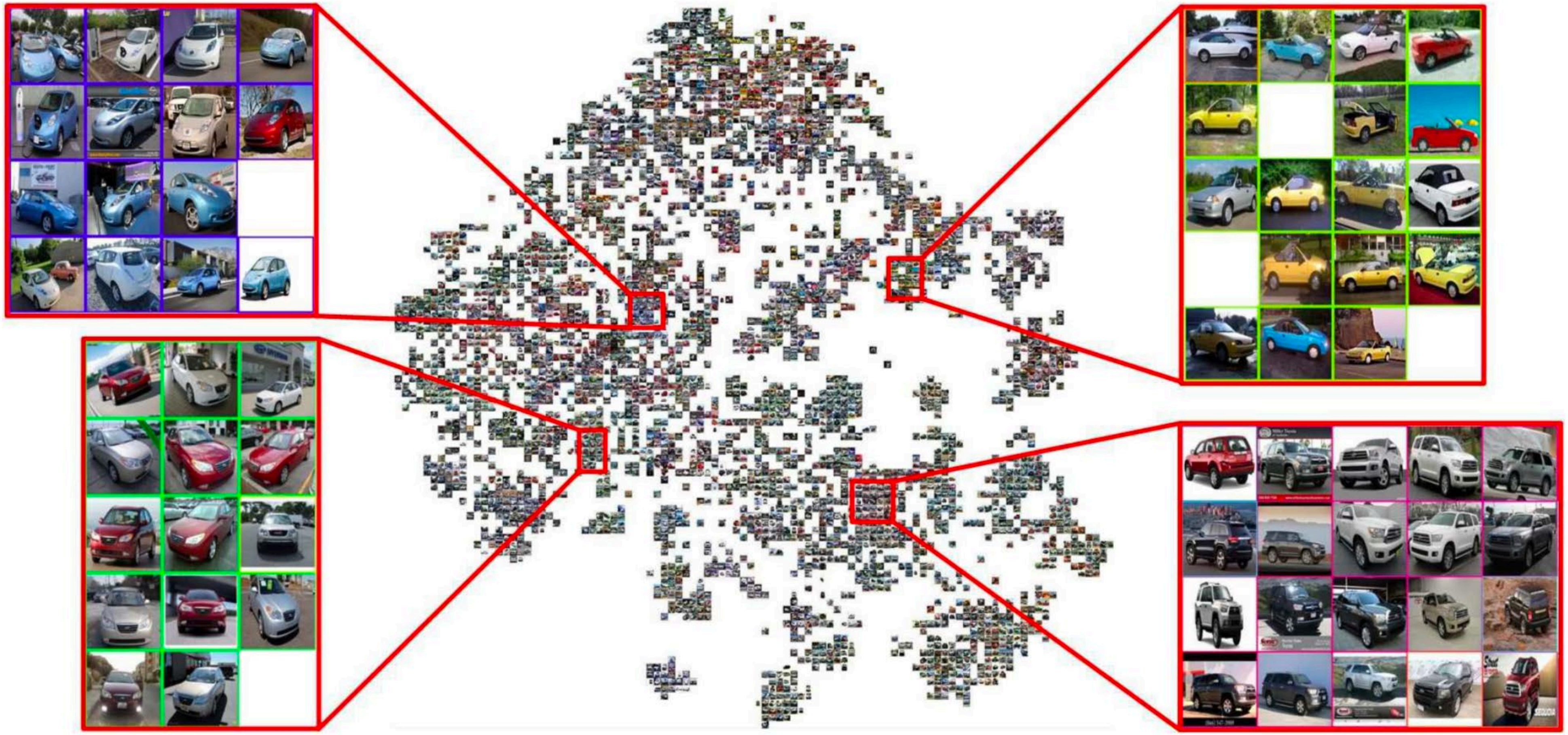


Figure 5. Visualization of the proposed DAML (N-pair) with Barnes-Hut t-SNE [37] on the Cars196 dataset, where the color of the border for each image represents the label. (Best viewed on a monitor when zoomed in.)

Experiment

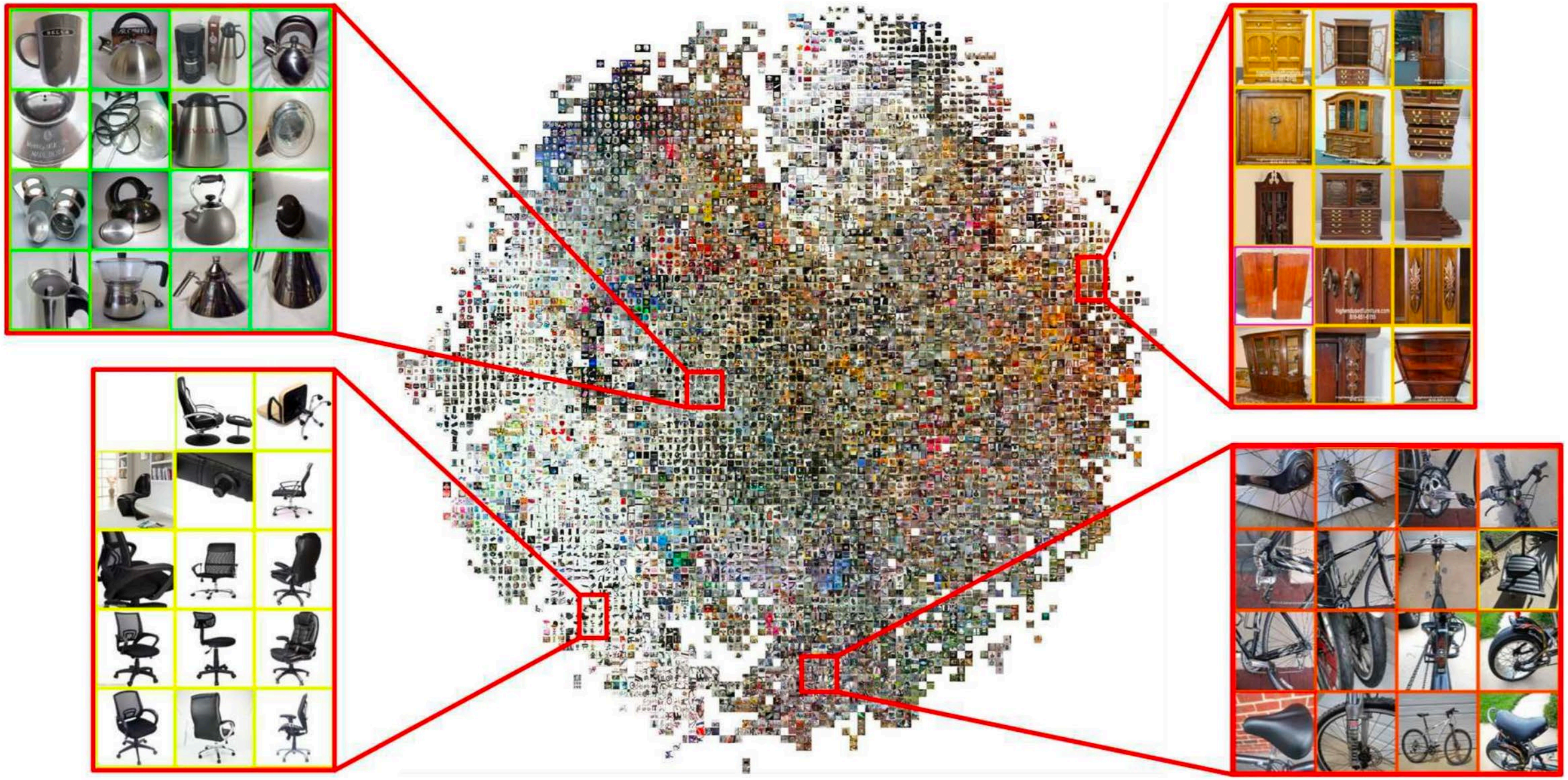


Figure 6. Visualization of the proposed DAML (N-pair) with Barnes-Hut t-SNE [37] on the Stanford Online Products dataset, where the color of the border for each image represents the label. (Best viewed on a monitor when zoomed in.)

4 Analize

Experiment

- **Achieved SOTA for the almost every metric learning task**

Table 1. Experimental results (%) on the CUB-200-2011 dataset compared with baseline methods.

Method	NMI	F_1	R@1	R@2	R@4	R@8
DDML	47.3	13.1	31.2	41.6	54.7	67.1
Triplet+N-pair	54.1	20.0	42.8	54.9	66.2	77.6
Angular	61.0	30.2	53.6	65.0	75.3	83.7
Contrastive	47.2	12.5	27.2	36.3	49.8	62.1
DAML (cont)	**49.1**	**16.2**	**35.7**	**48.4**	**60.8**	**73.6**
Triplet	49.8	15.0	35.9	47.7	59.1	70.0
DAML (tri)	**51.3**	**17.6**	**37.6**	**49.3**	**61.3**	**74.4**
Lifted	56.4	22.6	46.9	59.8	71.2	81.5
DAML (lifted)	**59.5**	**26.6**	**49.0**	**62.2**	**73.7**	**83.3**
N-pair	60.2	28.2	51.9	64.3	74.9	83.2
DAML (N-pair)	**61.3**	**29.5**	**52.7**	**65.4**	**75.5**	**84.3**

Table 2. Experimental results (%) on the Cars196 dataset compared with baseline methods.

Method	NMI	F_1	R@1	R@2	R@4	R@8
DDML	41.7	10.9	32.7	43.9	56.5	68.8
Triplet+N-pair	54.3	19.6	46.3	59.9	71.4	81.3
Angular	62.4	31.8	71.3	80.7	87.0	91.8
Contrastive	42.3	10.5	27.6	38.3	51.0	63.9
DAML (cont)	**42.6**	**11.4**	**37.2**	**49.6**	**61.8**	**73.3**
Triplet	52.9	17.9	45.1	57.4	69.7	79.2
DAML (tri)	**56.5**	**22.9**	**60.6**	**72.5**	**82.5**	**89.9**
Lifted	57.8	25.1	59.9	70.4	79.6	87.0
DAML (lifted)	**63.1**	**31.9**	**72.5**	**82.1**	**88.5**	**92.9**
N-pair	62.7	31.8	68.9	78.9	85.8	90.9
DAML (N-pair)	**66.0**	**36.4**	**75.1**	**83.8**	**89.7**	**93.5**

Table 3. Experimental results (%) on the Stanford Online Products dataset compared with baseline methods.

Method	NMI	F_1	R@1	R@10	R@100
DDML	83.4	10.7	42.1	57.8	73.7
Triplet+N-pair	86.4	21.0	58.1	76.0	89.1
Angular	87.8	26.5	67.9	83.2	92.2
Contrastive	82.4	10.1	37.5	53.9	71.0
DAML (cont)	**83.5**	**10.9**	**41.7**	**57.5**	**73.5**
Triplet	86.3	20.2	53.9	72.1	85.7
DAML (tri)	**87.1**	**22.3**	**58.1**	**75.0**	**88.0**
Lifted	87.2	25.3	62.6	80.9	91.2
DAML (lifted)	**89.1**	**31.7**	**66.3**	**82.8**	**92.5**
N-pair	87.9	27.1	66.4	82.9	92.1
DAML (N-pair)	**89.4**	**32.4**	**68.4**	**83.5**	**92.3**

Experiment

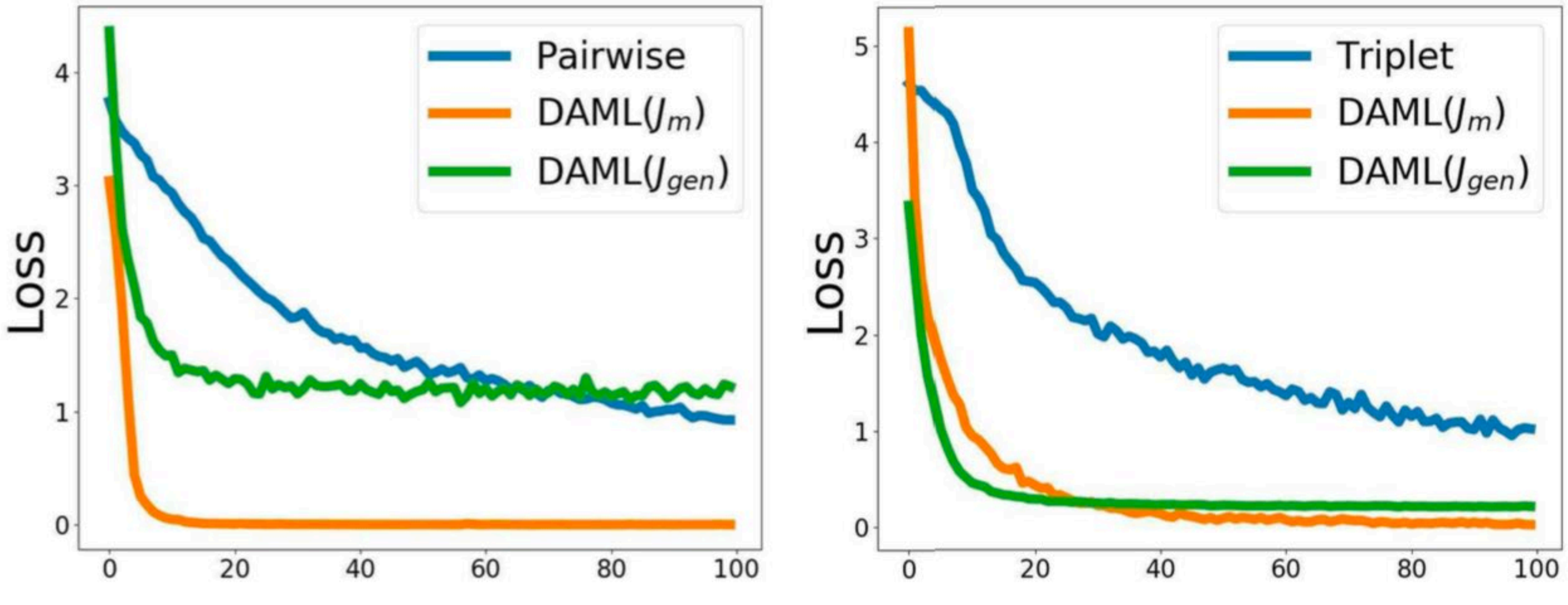
- **NMI: normalized mutual information**
- **F1: harmonic mean of the precision and recall**
- **R@K: Existence ratio of the Positive data in K nearest point.**

Table 1. Experimental results (%) on the CUB-200-2011 dataset compared with baseline methods.

Method	NMI	F_1	R@1	R@2	R@4	R@8
DDML	47.3	13.1	31.2	41.6	54.7	67.1
Triplet+N-pair	54.1	20.0	42.8	54.9	66.2	77.6
Angular	61.0	30.2	53.6	65.0	75.3	83.7
Contrastive	47.2	12.5	27.2	36.3	49.8	62.1
DAML (cont)	**49.1**	**16.2**	**35.7**	**48.4**	**60.8**	**73.6**
Triplet	49.8	15.0	35.9	47.7	59.1	70.0
DAML (tri)	**51.3**	**17.6**	**37.6**	**49.3**	**61.3**	**74.4**
Lifted	56.4	22.6	46.9	59.8	71.2	81.5
DAML (lifted)	**59.5**	**26.6**	**49.0**	**62.2**	**73.7**	**83.3**
N-pair	60.2	28.2	51.9	64.3	74.9	83.2
DAML (N-pair)	**61.3**	**29.5**	**52.7**	**65.4**	**75.5**	**84.3**

Experiment

Fastest converge



(a) Pairwise loss (b) Triplet loss

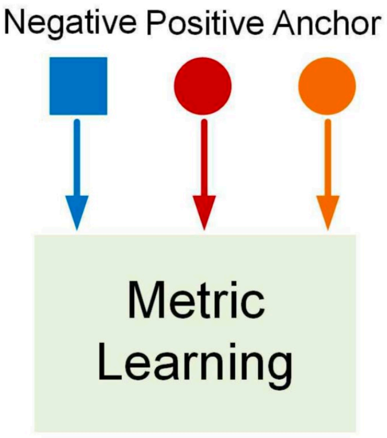
5 Discussion

Contribution

Practice

Contribution

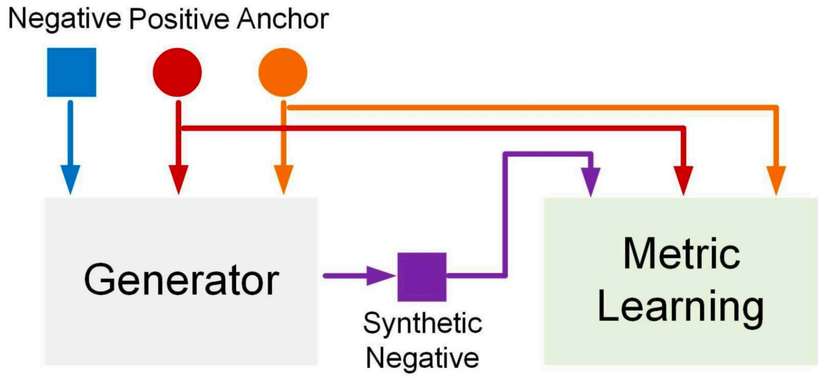
- **Hard Negative Mining**
  - Existing samples
- **Data Augmentation**
  - Fixed transformation
- **DAML**
  Create the Negative sample
  Various transformation



Embedding

$$\theta_f = \arg\min_{\theta_f} J_{\mathrm{m}}(\theta_f; \mathbf{x}_i, \mathbf{x}_i^+, \mathbf{x}_i^-, f)$$

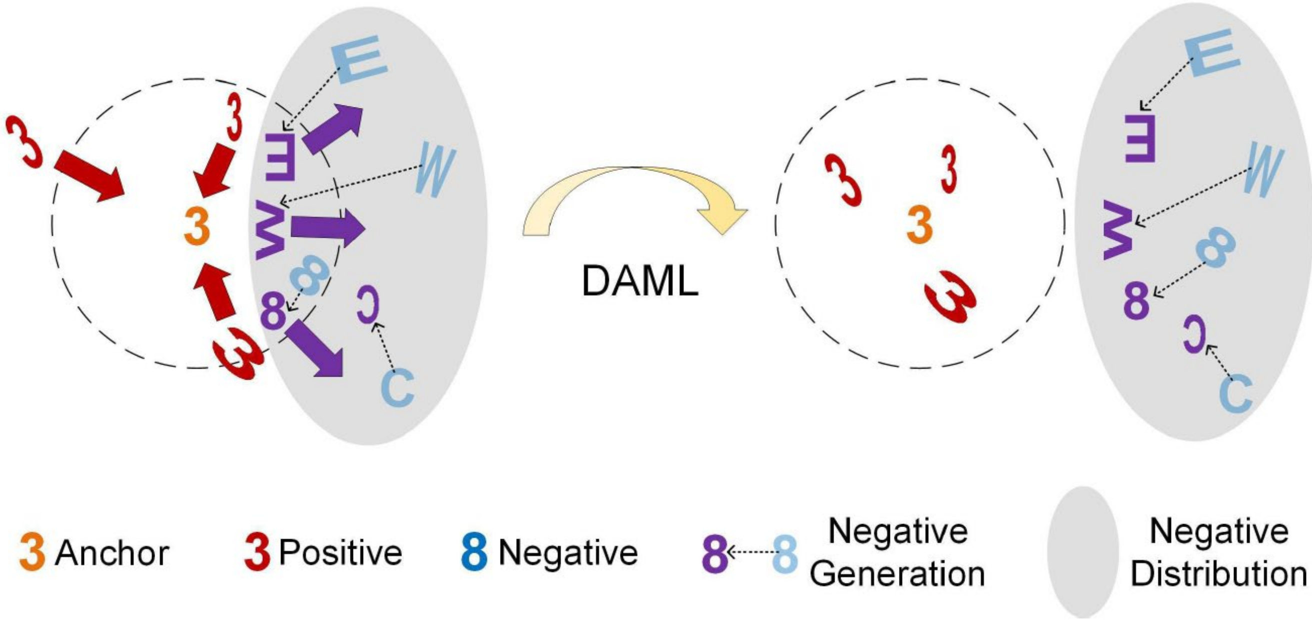
Weight

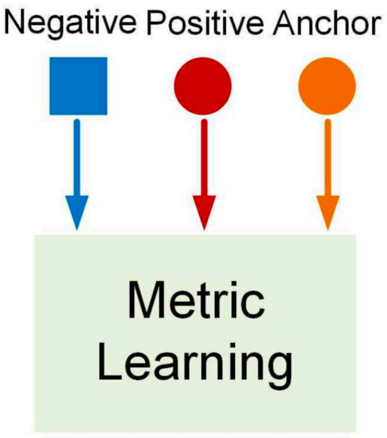


$$\theta_f^a = \arg\min_{\theta_f} J_{\mathrm{m}}(\theta_f; \mathbf{x}_i, \mathbf{x}_i^+, \widetilde{\mathbf{x}}_i^-, f)$$

Practice

- Paper assert that W, E could be a threaten after some transformation.

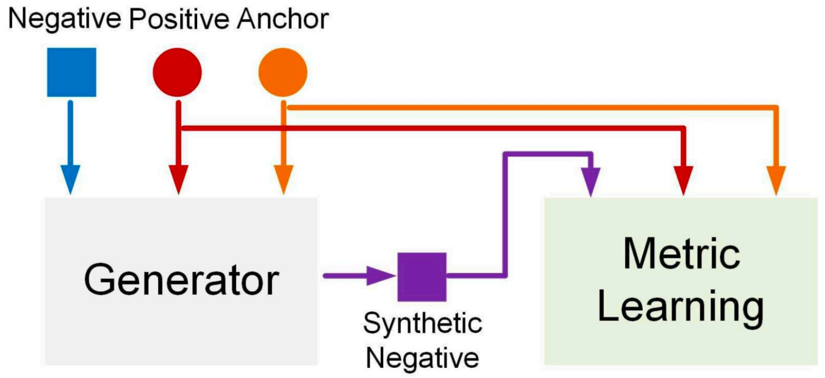




Embedding

$$\theta_f = \arg\min_{\theta_f} J_{\mathrm{m}}(\theta_f; \mathbf{x}_i, \mathbf{x}_i^+, \mathbf{x}_i^-, f)$$

Weight

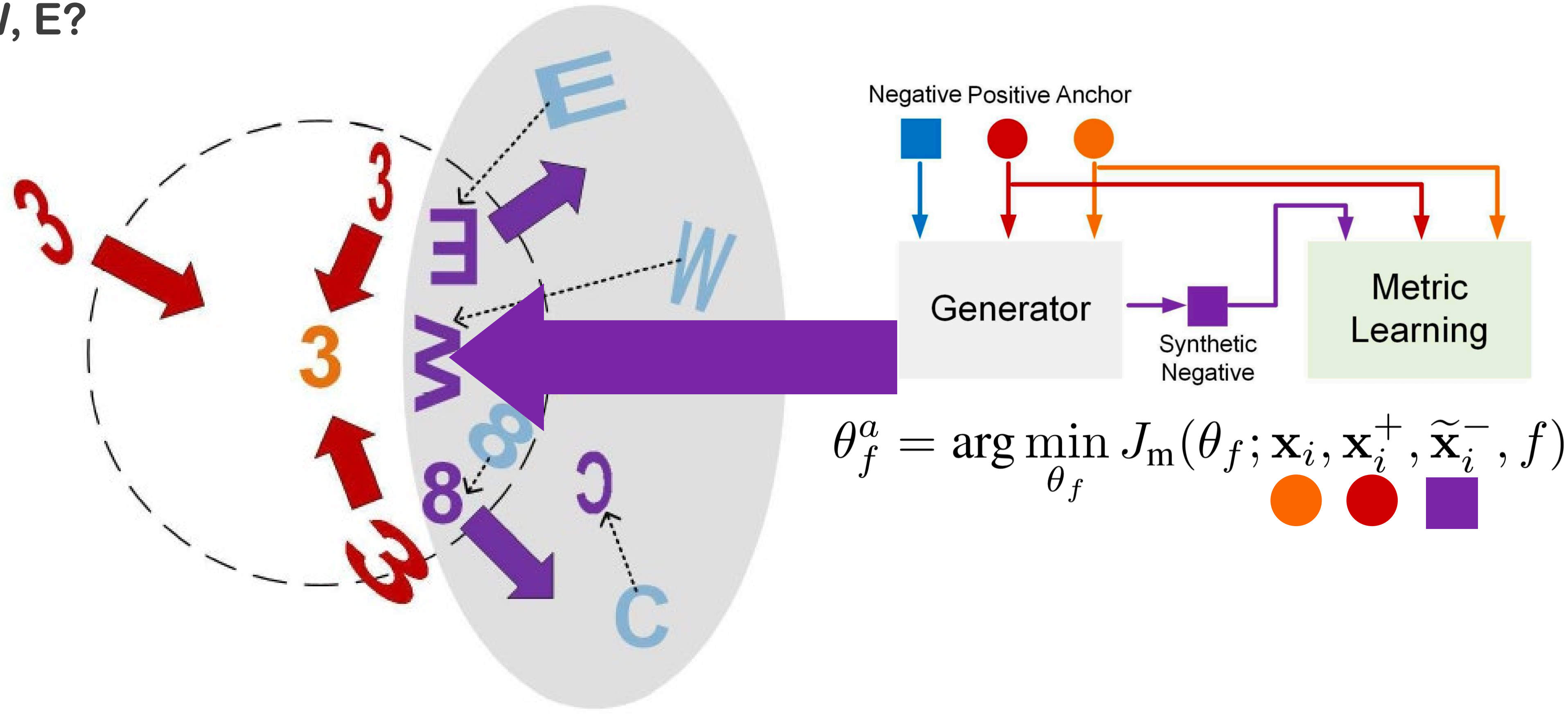


$$\theta_f^a = \arg\min_{\theta_f} J_{\mathrm{m}}(\theta_f; \mathbf{x}_i, \mathbf{x}_i^+, \widetilde{\mathbf{x}}_i^-, f)$$

Practice

- But the Generator really can make ʍ,Ǝ from the W, E?

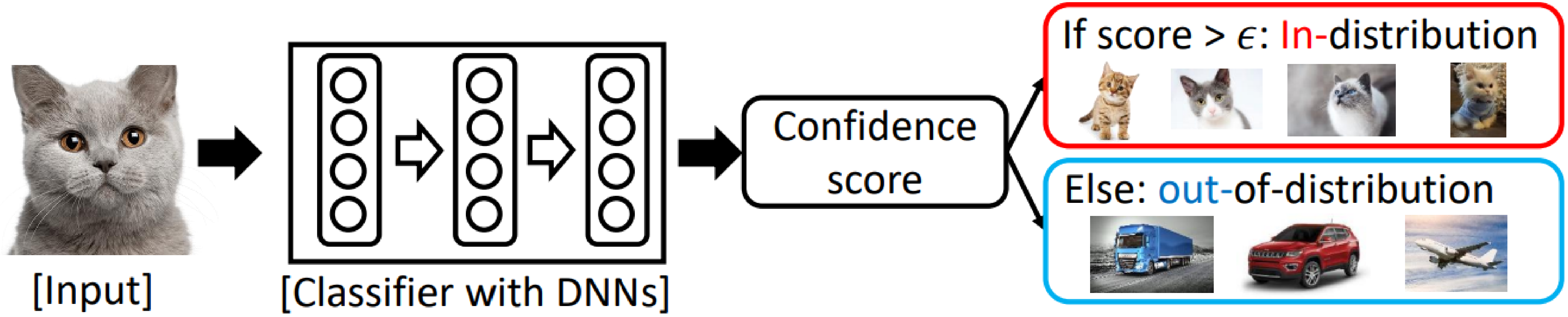
No



$$\theta_f^a = \arg\min_{\theta_f} J_{\mathrm{m}}(\theta_f; \mathbf{x}_i, \mathbf{x}_i^+, \widetilde{\mathbf{x}}_i^-, f)$$

Practice

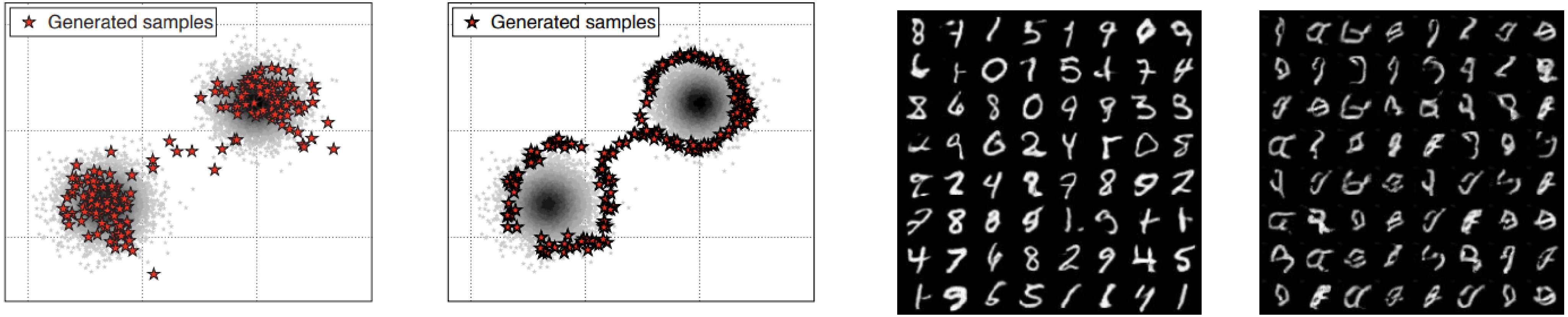
- Generated Adversarial example does not look a real, but seems like a mixture of the real image.



Practice

- Generated Adversarial example does not look a real, but seems like a mixture of the real image.

Comparison of original GAN and proposed GAN:



(a) Samples from the original GAN (b) Samples from the proposed GAN (c) Images from the original GAN (d) Images from the proposed GAN

Improved Deep Metric Learning with Multi-class N-pair Loss Objective, NIPS 2016

Practice

1. How did the DAML feed the Negative example?
   A. Generate the adversarial example
   B. Hard negative mining
   C. Over sampling

2. What was the best loss for the metric learning at the experiment?
   A. N-Pair Loss
   B. Contrastive Loss
   C. Triplet Loss

Results

Table 1. Experimental results (%) on the CUB-200-2011 dataset compared with baseline methods.

Method	NMI	F_1	R@1	R@2	R@4	R@8
DDML	47.3	13.1	31.2	41.6	54.7	67.1
Triplet+N-pair	54.1	20.0	42.8	54.9	66.2	77.6
Angular	61.0	30.2	53.6	65.0	75.3	83.7
Contrastive	47.2	12.5	27.2	36.3	49.8	62.1
DAML (cont)	**49.1**	**16.2**	**35.7**	**48.4**	**60.8**	**73.6**
Triplet	49.8	15.0	35.9	47.7	59.1	70.0
DAML (tri)	**51.3**	**17.6**	**37.6**	**49.3**	**61.3**	**74.4**
Lifted	56.4	22.6	46.9	59.8	71.2	81.5
DAML (lifted)	**59.5**	**26.6**	**49.0**	**62.2**	**73.7**	**83.3**
N-pair	60.2	28.2	51.9	64.3	74.9	83.2
DAML (N-pair)	**61.3**	**29.5**	**52.7**	**65.4**	**75.5**	**84.3**

Table 2. Experimental results (%) on the Cars196 dataset compared with baseline methods.

Method	NMI	F_1	R@1	R@2	R@4	R@8
DDML	41.7	10.9	32.7	43.9	56.5	68.8
Triplet+N-pair	54.3	19.6	46.3	59.9	71.4	81.3
Angular	62.4	31.8	71.3	80.7	87.0	91.8
Contrastive	42.3	10.5	27.6	38.3	51.0	63.9
DAML (cont)	**42.6**	**11.4**	**37.2**	**49.6**	**61.8**	**73.3**
Triplet	52.9	17.9	45.1	57.4	69.7	79.2
DAML (tri)	**56.5**	**22.9**	**60.6**	**72.5**	**82.5**	**89.9**
Lifted	57.8	25.1	59.9	70.4	79.6	87.0
DAML (lifted)	**63.1**	**31.9**	**72.5**	**82.1**	**88.5**	**92.9**
N-pair	62.7	31.8	68.9	78.9	85.8	90.9
DAML (N-pair)	**66.0**	**36.4**	**75.1**	**83.8**	**89.7**	**93.5**

Table 3. Experimental results (%) on the Stanford Online Products dataset compared with baseline methods.

Method	NMI	F_1	R@1	R@10	R@100
DDML	83.4	10.7	42.1	57.8	73.7
Triplet+N-pair	86.4	21.0	58.1	76.0	89.1
Angular	87.8	26.5	67.9	83.2	92.2
Contrastive	82.4	10.1	37.5	53.9	71.0
DAML (cont)	**83.5**	**10.9**	**41.7**	**57.5**	**73.5**
Triplet	86.3	20.2	53.9	72.1	85.7
DAML (tri)	**87.1**	**22.3**	**58.1**	**75.0**	**88.0**
Lifted	87.2	25.3	62.6	80.9	91.2
DAML (lifted)	**89.1**	**31.7**	**66.3**	**82.8**	**92.5**
N-pair	87.9	27.1	66.4	82.9	92.1
DAML (N-pair)	**89.4**	**32.4**	**68.4**	**83.5**	**92.3**

Results

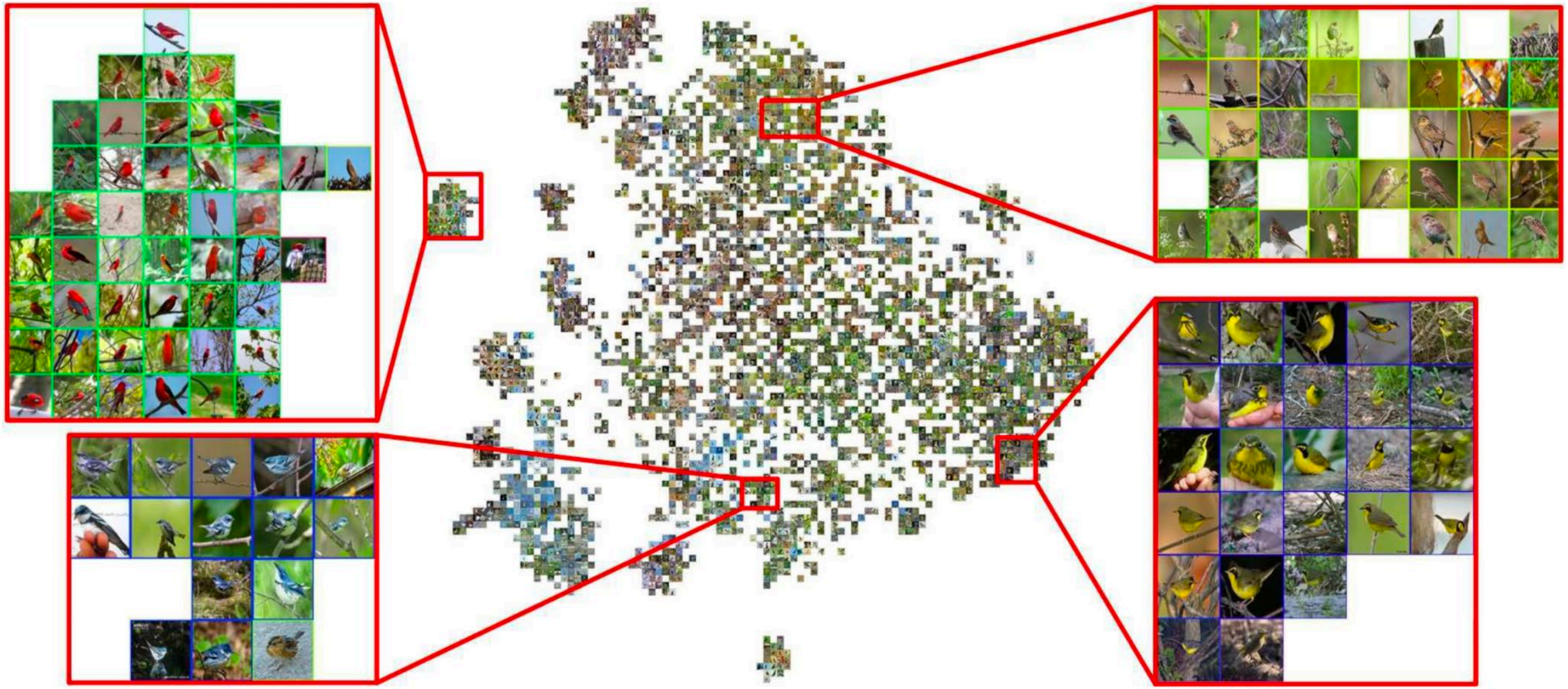


Figure 4. Visualization of the proposed DAML (N-pair) with Barnes-Hut t-SNE [37] on the CUB-200-2011 dataset, where the color of the border for each image represents the label. (Best viewed on a monitor when zoomed in.)

Results

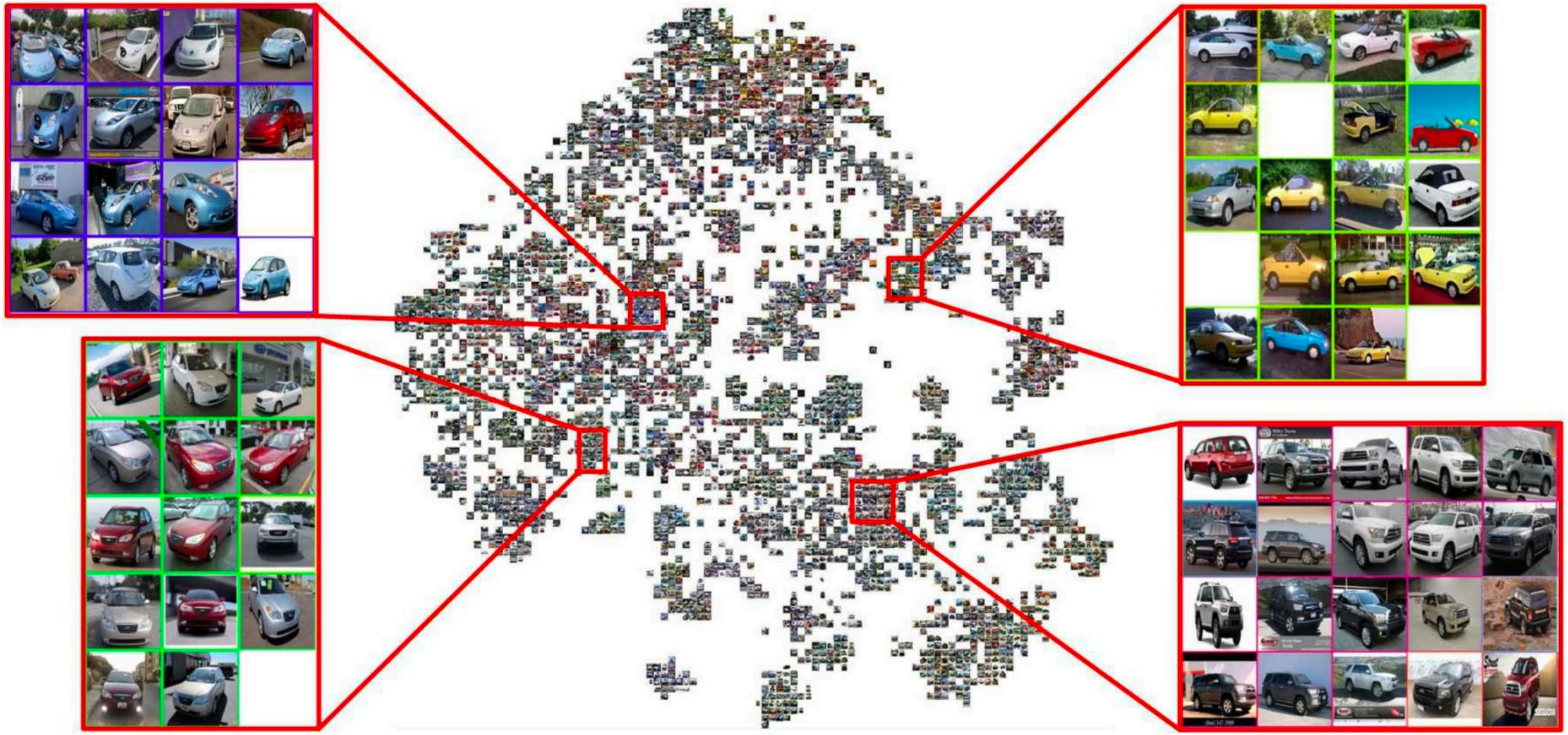


Figure 5. Visualization of the proposed DAML (N-pair) with Barnes-Hut t-SNE [37] on the Cars196 dataset, where the color of the border for each image represents the label. (Best viewed on a monitor when zoomed in.)

Results

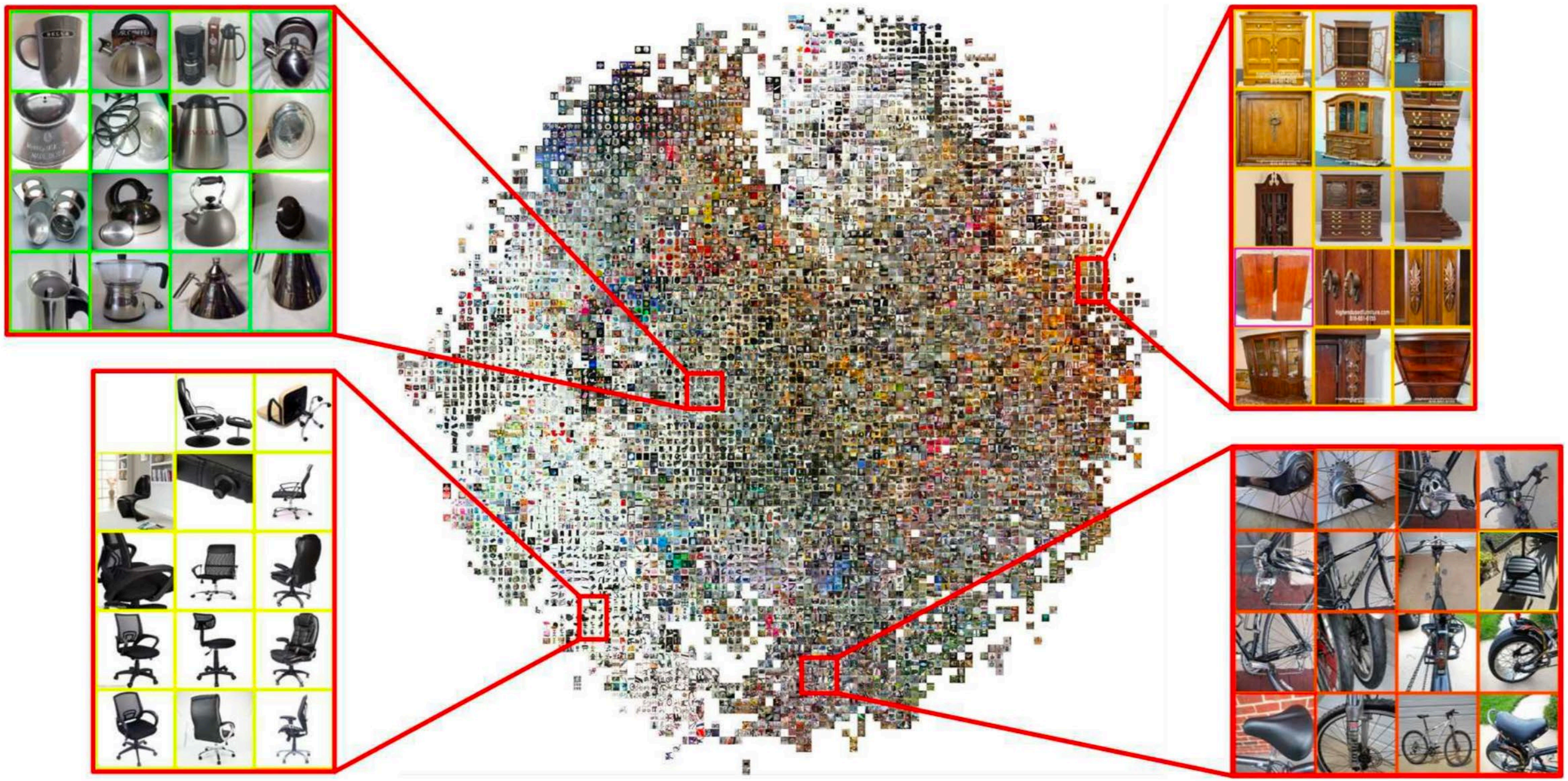
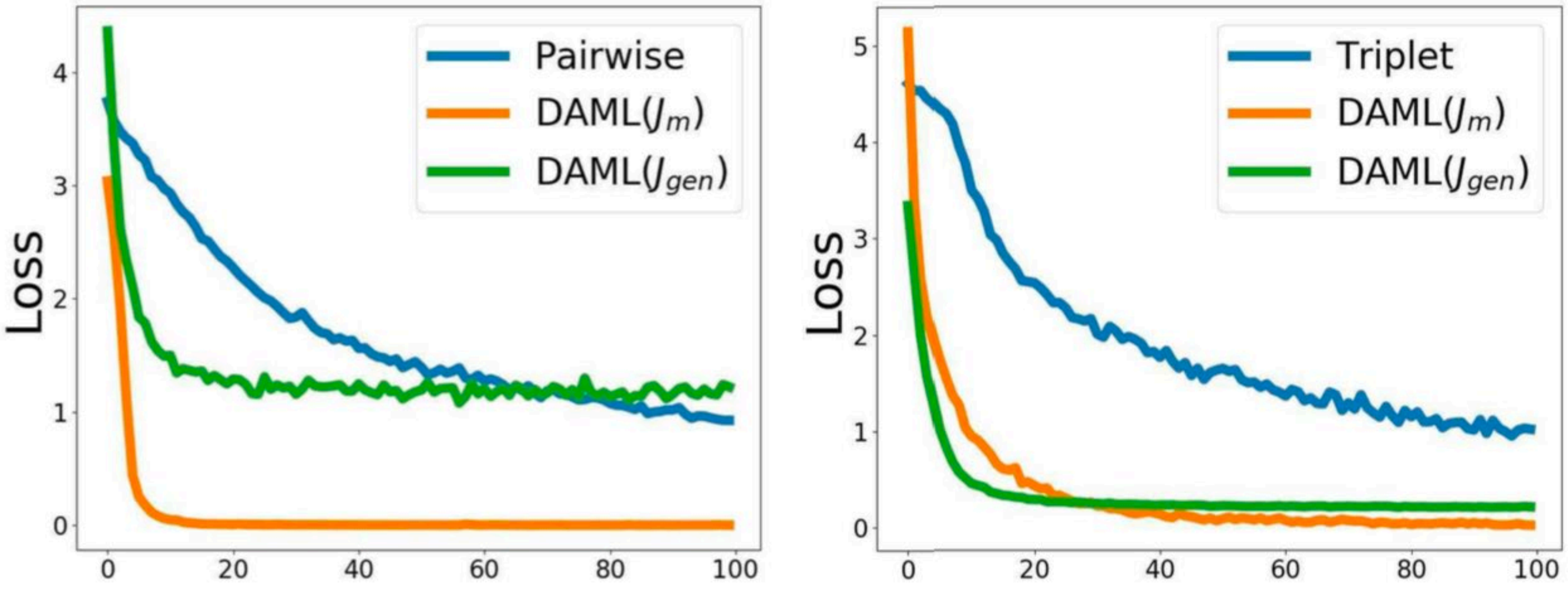


Figure 6. Visualization of the proposed DAML (N-pair) with Barnes-Hut t-SNE [37] on the Stanford Online Products dataset, where the color of the border for each image represents the label. (Best viewed on a monitor when zoomed in.)

Results



(a) Pairwise loss (b) Triplet loss

Algorithm 1: DAML

Input: Training image set, parameters λ, λ_1 and λ_2, margin α, and iteration numbers T.

Output: Parameters of the hard negative generator θ_g, and parameters of the metric function θ_f.

```
Pre-train θ_f without the hard negative generator.
Initialize θ_g.
for iter = 1, 2, ··· , T do
    Sample minibatch of m training images.
    Produce triplet or pairwise inputs from the batch.
    Jointly optimize θ_g and θ_f using (7).
end for
return θ_g and θ_f.
```