

Better and Faster: Exponential Loss for Image Patch Matching

ICCV2019 論文読み会

2019/12/11

株式会社エクサウィザーズ

MLエンジニア

小野 晃司

Contents

- 自己紹介
- 論文の概要
- タスクについて
- 関連研究
- 方法について
- 実験結果

自己紹介

- ・株式会社エクサウィザーズ 2年目
- ・生物系出身
- ・GNN (Graph Neural Network) などを中心の業務

論文の概要

論文の概要

- ・タスク：パッチマッチング
- ・貢献
 - ・新しいロスを提案
 - ・新しいサンプリング方法を提案
 - ・上記ロスに合わせて新しいネットワークを提案
- ・結果：UBCデータセットで最高精度を到達

タスクについて

タスク：パッチマッチングとは？



(a) Image 1



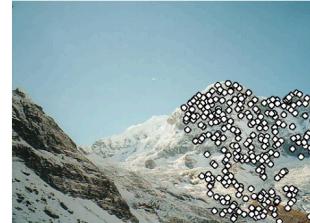
(b) Image 2



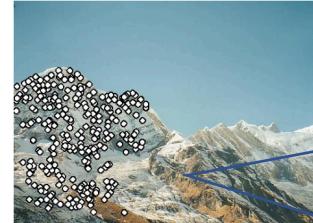
(c) SIFT matches 1



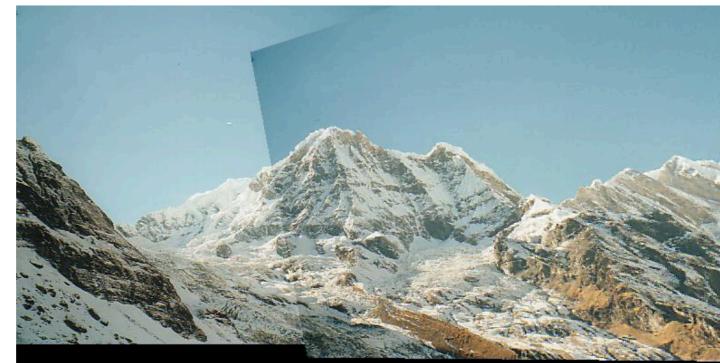
(d) SIFT matches 2



(e) RANSAC inliers 1



(f) RANSAC inliers 2



(g) Images aligned according to a homography

写真中の白い点（○）は写真の特徴。

特徴の抽出方法

- SIFT (古典的)
- SURF (古典的)

特徴（○）を使って写真を貼り合わせる。=>画像パッチマッチング

タスク：パッチマッチングとは？



(a) positive image pairs



(b) negative image pairs

Iaroslav Melekhov et al.(2016)

画像の貼り合わせだけではなく、
単純なシーンマッチングもパッチ
マッチング

負例のペアは自分たちで作
ることもできる。

パッチマッチングの難しいところ

- パッチの数が増えると…
 - 簡単なペアが多くなってしまう。
 - 似たようなペアが多くなってしまう。

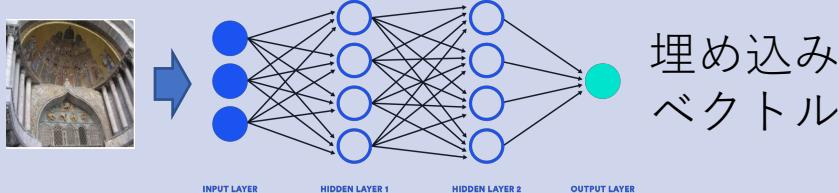
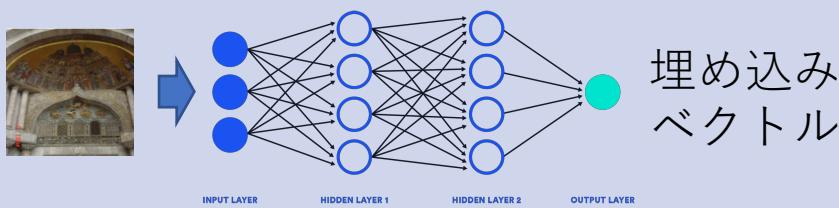


関連研究

Metric LearningとDescriptor Learning

Descriptor Learning

- モデルの出力のペアの距離を算出

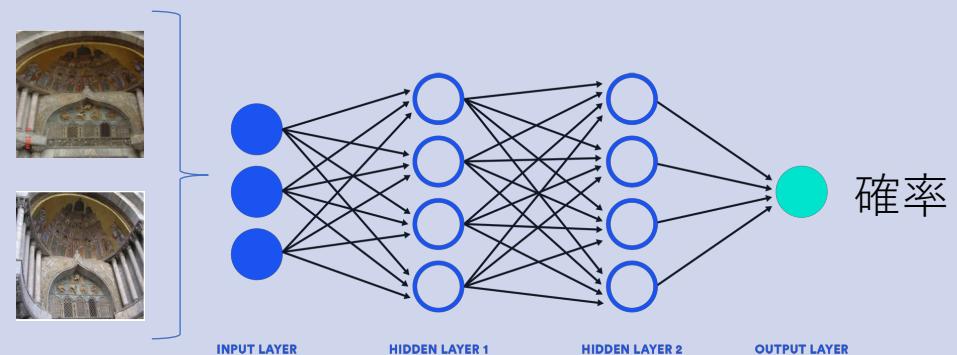


二つの埋め込みベクトルから距離を計算し、距離を最適化しネットワークを更新する



Metric Learning

- ペアのマッチング確率を算出



ペアを入力して、正例なら1を負例なら0であるように、ネットワークを更新する。



方法について

- Exponential Losses
- Hard positive and negative sample mining
- Shared feature network

Siamese Loss (Descriptor Learning)

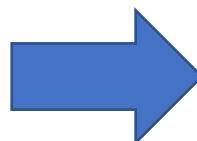
$$\mathcal{L}_{Siamese} = D_{ap} + [\alpha - D_{mn}]_+$$

(a, p) は正例、 (m, n) は負例

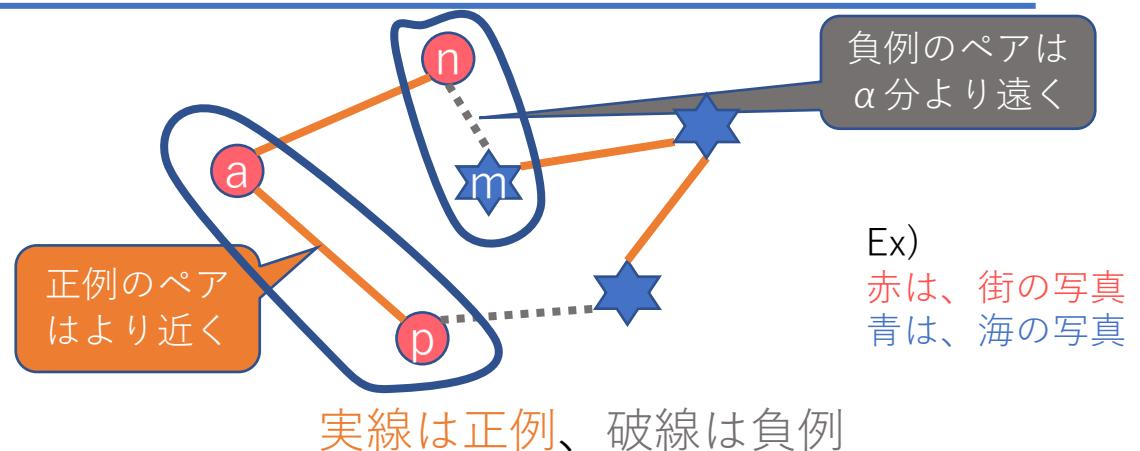
$$D_{ap} = \|f(a) - f(p)\|_2$$

$$[z]_+ = \max(z, 0)$$

$$D_{mn} = \|f(m) - f(n)\|_2$$



正例ペアはより近くに、一方で負例ペアは正例より α だけ離れるようにデザインされている。



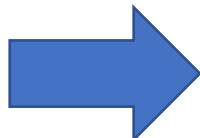
Siamese Loss (Descriptor Learning)

$$\mathcal{L}_{Siamese} = D_{ap} + [\alpha - D_{mn}]_+$$

(a, p) は正例、 (m, n) は負例 $D_{ap} = \|f(a) - f(p)\|_2$

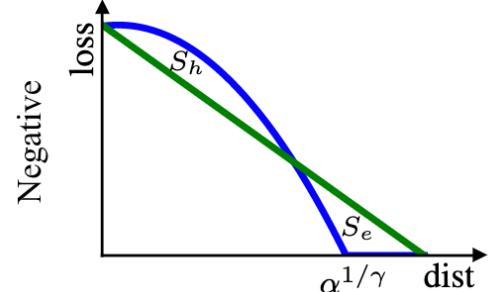
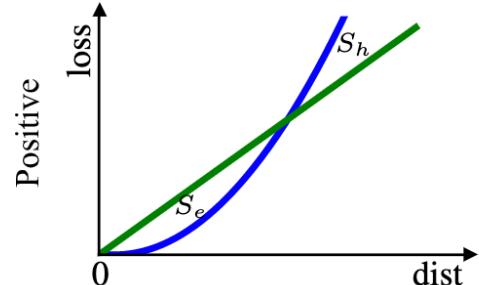
$$[z]_+ = \max(z, 0)$$

$$D_{mn} = \|f(m) - f(n)\|_2$$



正例ペアはより近くに、一方で負例ペアは正例より α だけ離れるようにデザインされている。

緑の線が正例、負例のロスへの影響



Triplet Loss (Descriptor Learning)

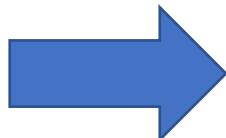
$$\mathcal{L}_{triplet} = [D_{ap} - D_{an} + \alpha]_+$$

(a, p) は正例、 (a, n) は負例

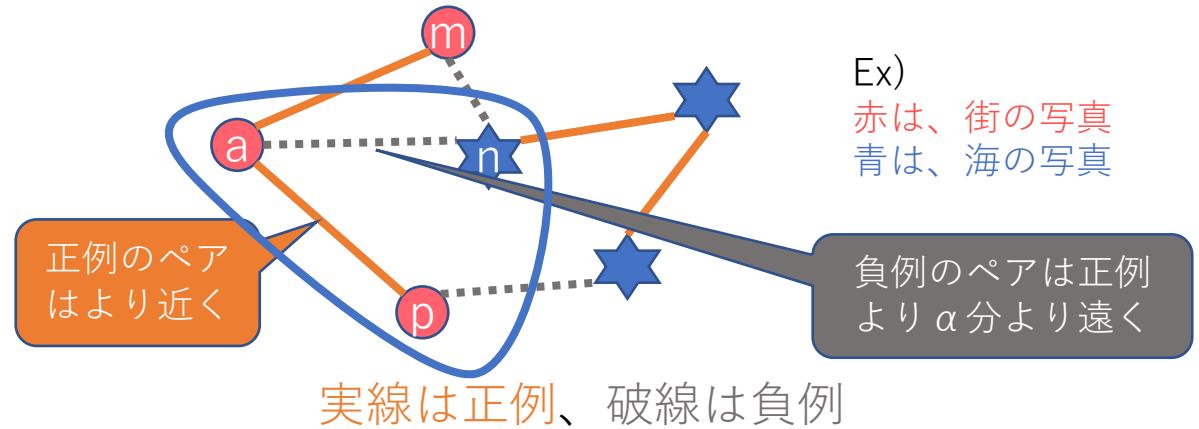
$$D_{ap} = \|f(a) - f(p)\|_2$$

$$[z]_+ = \max(z, 0)$$

$$D_{an} = \|f(a) - f(n)\|_2$$



負例ペアは正例ペアより α だけ大きくないといけない。負例ペアと正例ペアの関係を表しており、Siamese losses より収束が速い傾向にある。



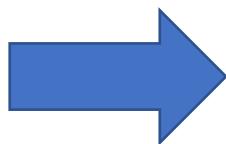
Triplet Loss (Descriptor Learning)

$$\mathcal{L}_{triplet} = [D_{ap} - D_{an} + \alpha]_+$$

(a, p) は正例、 (a, n) は負例 $D_{ap} = \|f(a) - f(p)\|_2$

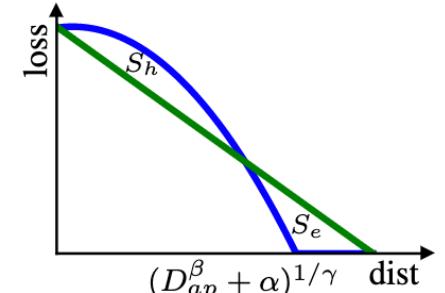
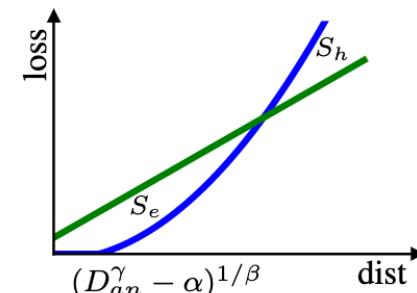
$$[z]_+ = \max(z, 0)$$

$$D_{an} = \|f(a) - f(n)\|_2$$



負例ペアは正例ペアより α だけ大きくないといけない。負例ペアと正例ペアの関係を表しており、Siamese losses より収束が速い傾向にある。

緑の線が正例、負例のロスへの影響



「提案手法」 Exponential Losses (Descriptor Learning)

$$\mathcal{L}_{Siamese}(dist) = D_{ap}^{\beta} + [\alpha - D_{mn}^{\gamma}]_+$$

$$\mathcal{L}_{triplet}(dist) = [D_{ap}^{\beta} - D_{an}^{\gamma} + \alpha]_+$$

(a, p) は正例、 (a, n) は負例

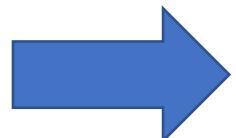
$$D_{ap} = \|f(a) - f(p)\|_2$$

(a, n) は負例

$$D_{mn} = \|f(a) - f(n)\|_2$$

$$[z]_+ = \max(z, 0) \quad \beta, \gamma > 0$$

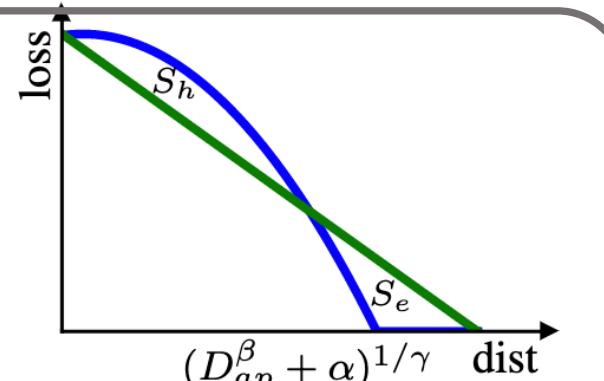
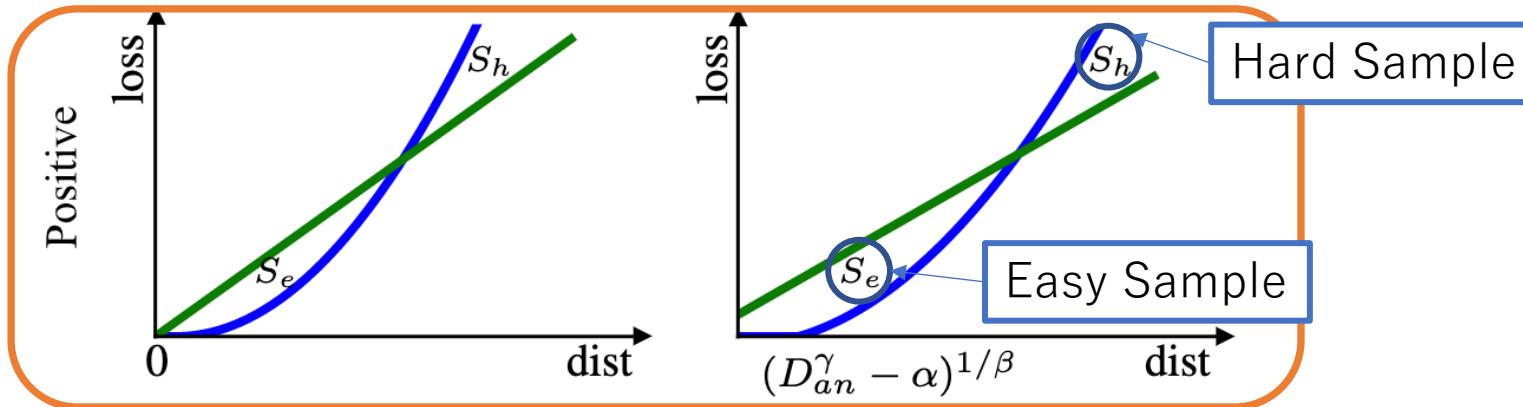
$$D_{an} = \|f(a) - f(n)\|_2$$



β, γ がそれぞれ 1 ならば、Siamese loss や Triplet loss になる。

「提案手法」 Exponential Losses (Descriptor Learning)

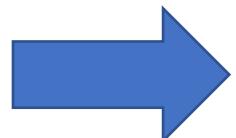
青の線が
Exponential
Lossに対応。



(a) Siamese loss (dist)

(b) triplet loss (dist)

Best parameters
 $\beta = 2, \gamma = 2$



Easy Sampleにはより鈍感に、 Hard Sampleにはより敏感に。

「提案手法」 Exponential Losses (Metric Learning)

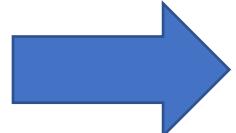
$$\mathcal{L}_{Siamese(prob)} = P_{ap}^{\beta} + [\alpha - P_{mn}^{\gamma}]_+$$

$$\mathcal{L}_{triplet(prob)} = [P_{ap}^{\beta} - P_{an}^{\gamma} + \alpha]_+$$

(a, p) は正例、 (m, n) は負例 P_{ap} : 正例の確率

(a, n) は負例 P_{mn} : 負例の確率

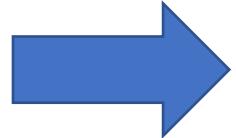
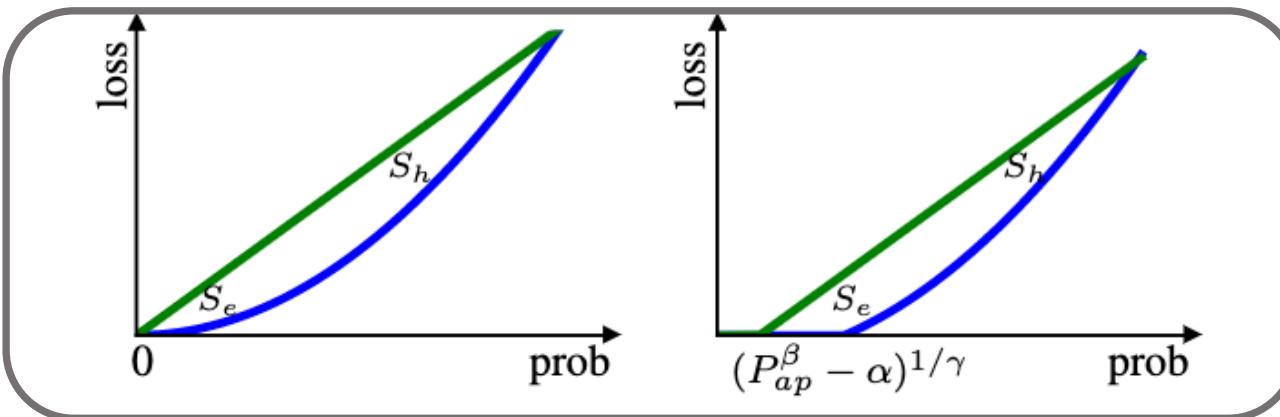
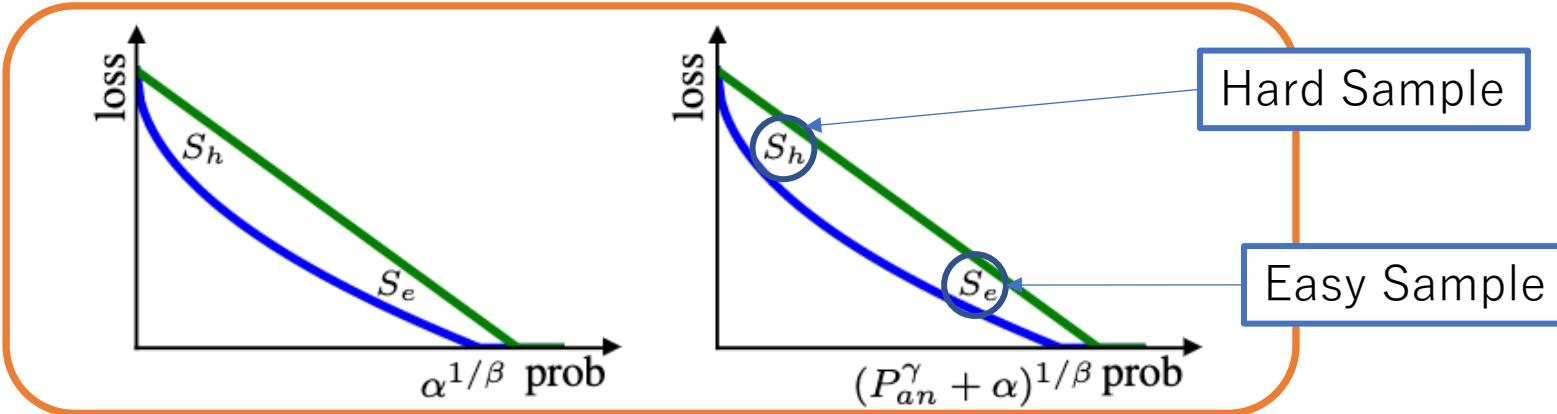
$[z]_+ = \max(z, 0)$ $\beta, \gamma > 0$ P_{an} : 負例の確率



β, γ がそれぞれ 1 ならば、Siamese loss や Triplet loss になる。

「提案手法」 Exponential Losses (Metric Learning)

青の線が
Exponential
Lossに対応。



Easy Sampleにはより鈍感に、 Hard Sampleにはより敏感に。

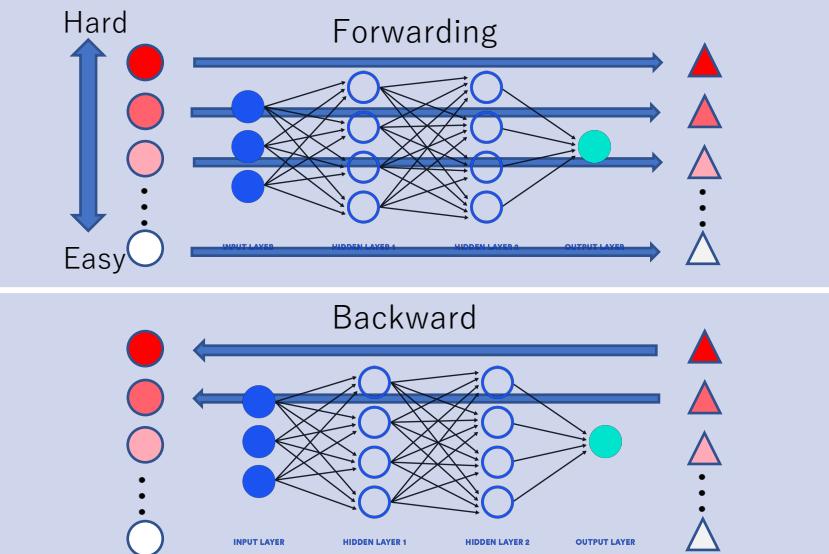
方法について

- Exponential Losses
- Hard positive and negative sample mining
- Shared feature network

Hard positive and Negative sample mining

- Lossだけを改善しても、Easy Samplesが多いとモデルの性能が向上しない。

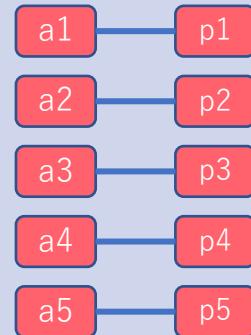
Hard Positive Samples Mining



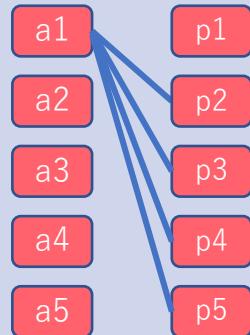
上位k個のHard Positive samplesだけを使ってパラメータをアップデートする。

Hard Negative Samples Mining

正例集合



負例集合



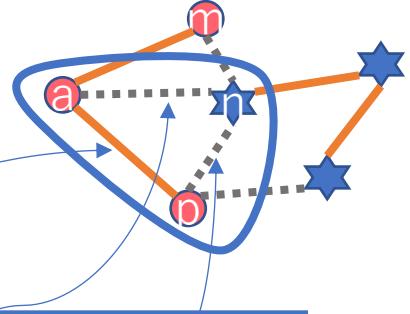
a2~a5も同様に負例を作る

正例を使って、 $n(n-1)$ の負例のクロススペアを作成する。

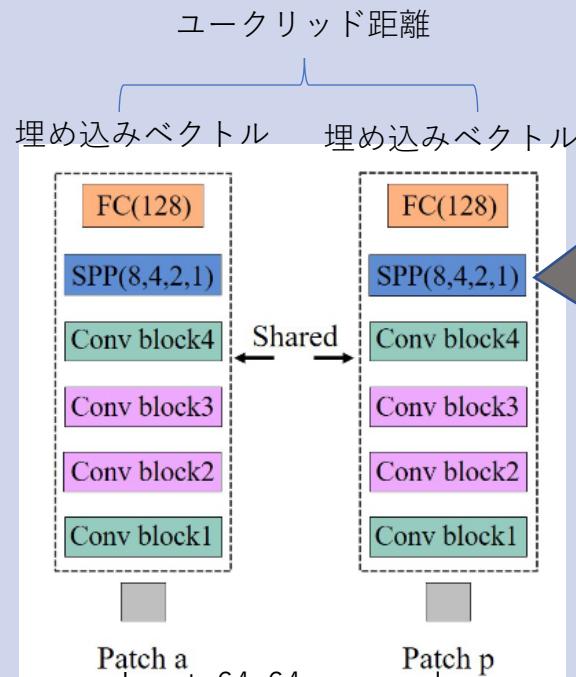
方法について

- Exponential Losses
- Hard positive and negative sample mining
- Shared feature network

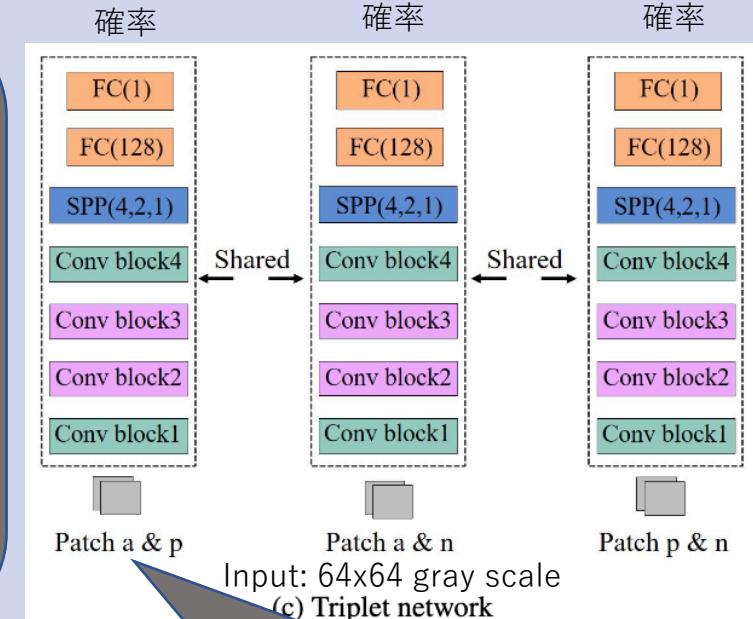
Shared feature network



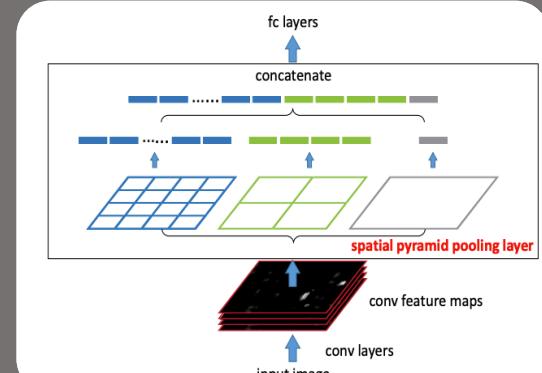
Descriptor Learning



Metric Learning



Spatial Pyramid Pooling



Inputは二つの画像をチャネル方向に重ね合わせる。

実験結果

Hypher parameter turning

$\beta \backslash \gamma$	0.1	0.3	0.5	1.0
1.0	1.05	0.90	1.16	1.40
2.0	1.14	0.75	1.12	0.95
3.0	1.01	0.78	0.79	0.81
4.0	1.55	0.91	0.78	0.79

(a) Exp-TLoss (prob) with varied β and γ .

$\beta \backslash \gamma$	0.5	1.0	2.0	3.0
0.5	1.97	1.09	2.24	6.23
1.0	2.22	1.37	1.12	2.00
2.0	1.15	2.23	1.03	1.11
3.0	1.11	1.99	1.22	1.07

(b) Exp-TLoss (dist) with varied β and γ .

$bs \backslash mr$	0:1	1:8	1:4	1:2	1:1
64	1.13	1.07	1.11	1.23	1.27
128	1.09	1.12	1.07	1.03	1.13
256	1.11	1.14	1.10	1.09	1.35
512	1.25	1.22	1.29	1.34	1.49

(c) FPR95 under varied bs and mr .

Table 1: Ablation study on varied exponential orders, β and γ , batch size, and mining ratio. (a) FPR95 achieved by Exp-TLoss with varied β and γ in metric learning; (b) FPR95 achieved by Exp-TLoss with varied β and γ in descriptor learning; (c) FPR95 achieved by varied batch size (bs) and mining ratio (mr) in descriptor learning.

- UBCデータセットからサブセットを作成し学習、他のサブセットで上記を評価
- FPR95：閾値0.95のときのFalse positive比率(小さい方がよい。)

UBC Dataset

評価

学習

Method	Feature Dim	NOT YOS		LIB	YOS	LIB	NOT	Mean	STD
		LIB	NOT						
Metric Learning									
SIFT[23]	128	29.84	22.5		27.29	26.55	3.04		
MatchNet[13]	4096	6.90	10.77	3.87	5.67	10.88	8.39	7.74	2.56
SNet-GLoss+[21]	256	6.39	8.43	1.84	2.83	6.61	5.57	5.27	2.27
DeepCompare 2ch-deep+[39]	256	4.55	7.40	2.01	2.52	4.75	4.38	4.27	1.75
DeepCompare 2ch-2stream+[39]	256	4.85	7.20	1.90	2.11	5.00	4.10	4.19	1.81
CS SNet-GLoss+[21]	384	3.69	4.91	0.77	1.14	3.09	2.67	2.71	1.42
Exp-TLoss (prob)+	128	0.44	1.07	1.63	1.85	3.78	2.27	1.84	1.04
Descriptor Learning									
TL+AS+GOR[40]	128	1.95	5.40	4.80	5.15	6.45	2.38	4.36	1.63
L2-Net[36]	128	3.64	5.29	1.15	1.62	4.43	3.30	3.23	1.46
CS L2-Net[36]	256	2.55	4.24	0.87	1.39	3.81	2.84	2.61	1.20
L2-Net+[36]	128	2.36	4.70	0.72	1.29	2.57	1.71	2.22	1.27
CS L2-Net+[36]	256	1.71	3.87	0.56	1.09	2.07	1.30	1.76	1.05
Scale aware[19]	128	0.68	2.51	1.79	1.64	2.96	1.02	1.64	0.79
HardNet+[27]	128	0.53	1.96	1.49	1.84	2.51	0.78	1.51	0.69
Exp-TLoss (dist)	128	0.47	1.32	1.16	1.10	2.01	0.67	1.12	0.49

Table 2: PFR95 comparisons among our proposal and state-of-the-arts on UBC Benchmark. + indicates data augmentation.
 LIB: Liberty, NOT: Notredame, YOS: Yosemite.

まとめ

- ・タスク：パッチマッチング
- ・貢献
 - ・新しいロスを提案=> **Exponential losses**
 - ・新しいサンプリング方法を提案 => **Hard positive and negative mining**
 - ・上記ロスに合わせて新しいネットワークを提案提案 => **Shared feature network**
- ・結果：UBCデータセットで最高精度を到達

