Classification is a Strong Baseline for Deep Metric Learning

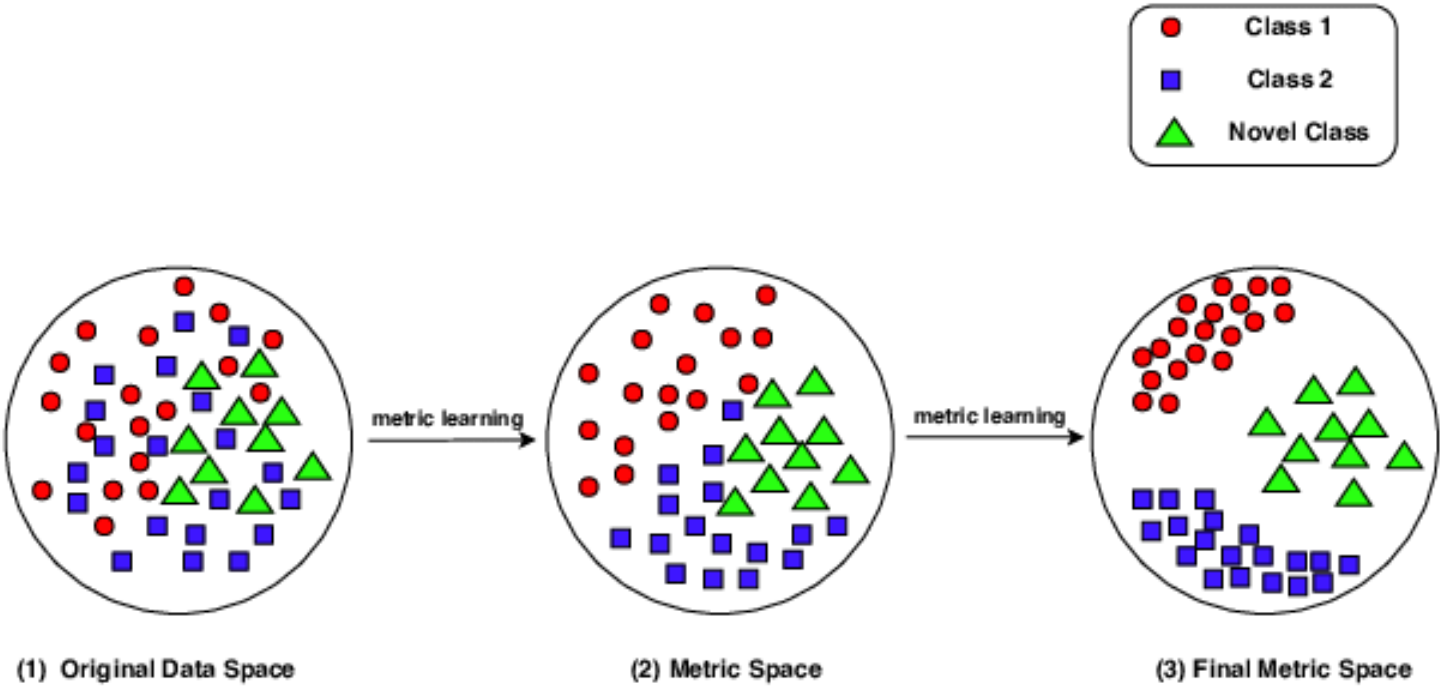
Andrew Zhai, Hao-Yu Wu

Presented by M.Rodin

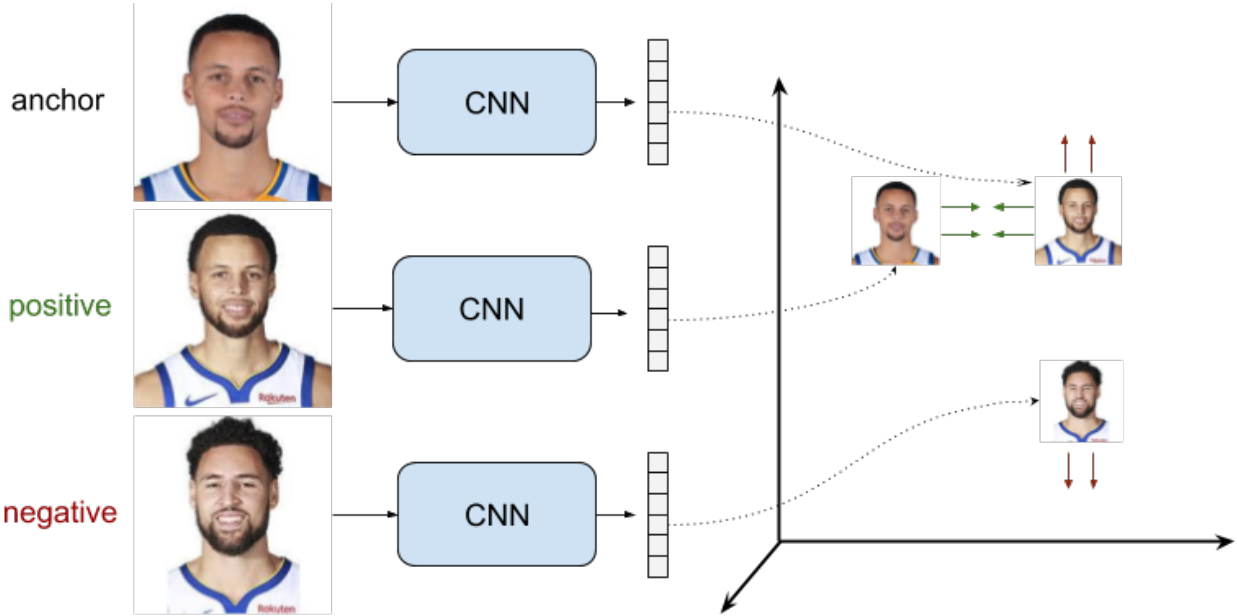
November 17, 2020

Outline

Problem formulation

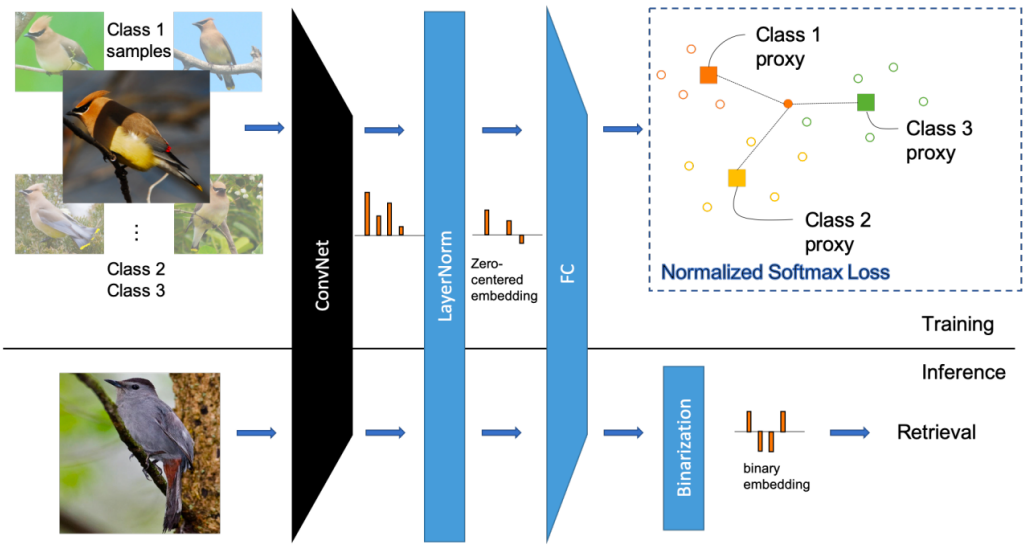


Triplet loss



$$L(r_a, r_p, r_n) = max(0, m + d(r_a, r_p) - d(r_a, r_n)) \quad (1)$$

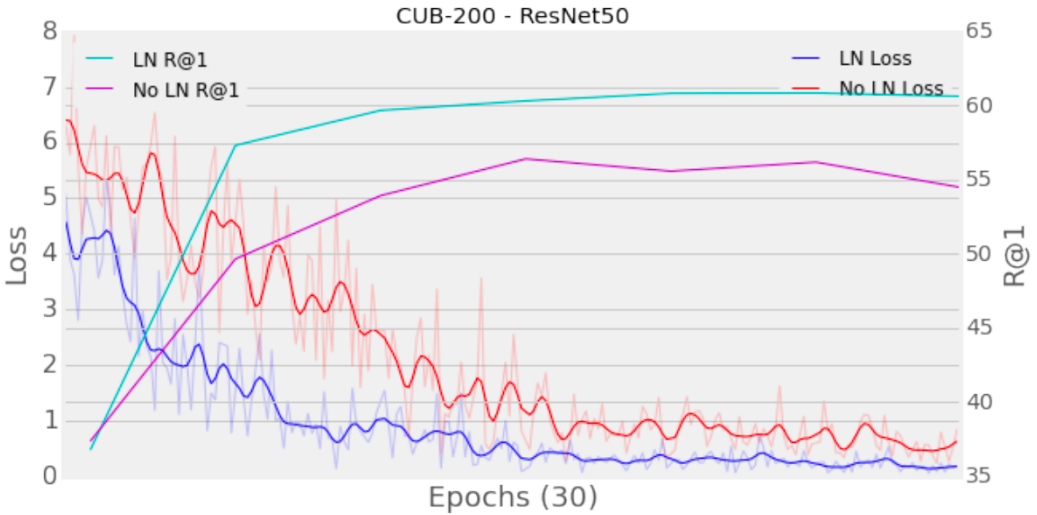
NormSoftmax: Architecture



$$L_{\text{norm}} = -\log\left(\frac{\exp\left(x^T p_y/\sigma\right)}{\sum_{z\in Z}\exp\left(x^T p_z/\sigma\right)}\right) \tag{2}$$

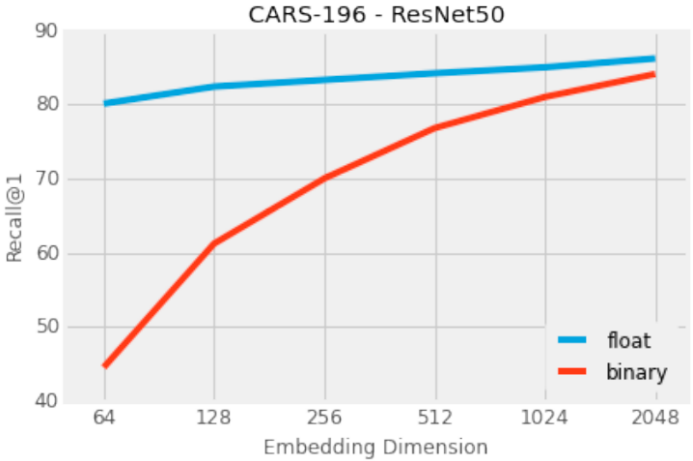
NormSoftmax: LayerNorm

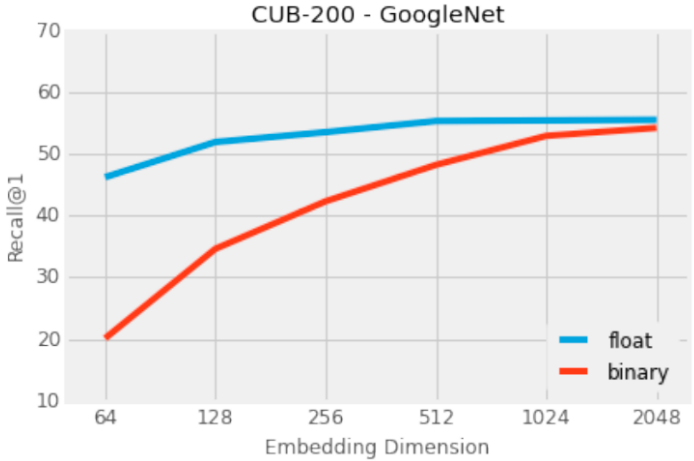
- Allows us to easily binarize embeddings via thresholding at zero
- Helps the network better initialize new parameters and reach better optima



NormSoftmax: Binarization

- thresholding at zero
- 2048 binary vector = 64 float vector (256 bytes in memory)





NormSoftmax: Batch construction

S	-	1	3	12	25	37	75
C	-	75	25	6	3	2	1
R@1	59.5	59.6	60.0	60.8	61.3	59.6	40.9

Table 4: ResNet50 Recall@1 on CUB-200-2011 dataset across varying samples per class for batch size of 75. (**S**) Samples per class in batch. (**C**) Distinct classes in batch. First column shows no class balancing in batch

NormSoftmax: Results

	Net.	CARS-196				CUB-200			
Recall@K		1	2	4	8	1	2	4	8
Contrastive^{128} [18]	G	21.7	32.3	46.1	58.9	26.4	37.7	49.8	62.3
Lifted Struct^{128} [18]	G	49.0	60.3	72.1	81.5	47.2	58.9	70.2	80.2
Clustering^{64} [19]	B	58.1	70.6	80.3	87.8	48.2	61.4	71.8	81.9
Npairs^{64} [17]	G	71.1	79.7	86.5	91.6	51.0	63.3	74.3	83.2
Angular Loss^{512} [9]	G	71.4	81.4	87.5	92.1	54.7	66.3	76.0	83.9
Proxy NCA^{64} [14]	B	73.2	82.4	86.4	88.7	49.2	61.9	67.9	72.4
HDC^{384} [30]	G	73.7	83.2	89.5	93.8	53.6	65.7	77.0	85.6
Margin^{128} [27]	R50	79.6	86.5	91.9	95.1	63.6	74.4	83.1	90.0
HTL^{512} [24]	B	81.4	88.0	92.7	95.7	57.1	68.8	78.7	86.5
A-BIER^{512} [15]	G	82.0	89.0	93.2	96.1	57.5	68.7	78.3	86.2
ABE-8^{512} [26]	G†	85.2	90.5	94.0	96.1	60.6	71.5	79.8	87.4
DREML^{576} [29]	R18	86.0	91.7	95.0	97.2	63.9	75.0	83.1	89.7
LMCL^{512} [23]	R50	73.9	81.7	87.4	91.5	58.7	70.3	79.9	86.9
LMCL^{*2048} [23]	R50	88.3	93.1	95.7	97.4	61.2	71.4	80.4	87.4
NormSoftMax^{1024}	B	87.9	93.2	96.2	98.1	62.2	73.9	82.7	89.4
NormSoftmax^{128}	R50	81.6	88.7	93.4	96.3	56.5	69.6	79.9	87.6
NormSoftmax^{512}	R50	84.2	90.4	94.4	96.9	61.3	73.9	83.5	90.0
NormSoftmax^{2048}	R50	**89.3**	**94.1**	**96.4**	**98.0**	**65.3**	**76.7**	**85.4**	**91.8**
NormSoftmax$^{2048\text{bits}}$	R50	88.7	93.7	**96.4**	**98.0**	63.3	75.2	84.3	91.0

Conclusion

- Classification-based metric learning approaches can achieve state-of-the-art
- Binarization is allowing us to achieve SOTA performance with the same memory footprint as 64 dimensional float embeddings

Practice

Tutorial in the similarity section
`https://github.com/microsoft/computervision-recipes/`

Practice: Document stamps detection

Thank you for your attention