Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere

School of Industrial and Management Engineering, Korea University

Sae Rin Lim

• Contents

• Introduction

❖ **Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere**

- 2020 ICML에 게재, 2022년 06월 09일 기준 377회 인용
- Contrastive learning의 표현학습에 대한 새로운 insight를 제공하는 논문
- Contrastive loss의 역할을 Unit hypersphere상에서 학습되는 normalized representation vector들의 Alignment와 Uniformity의 관점에서 해석(contrastive learning을 설명할 때 많이 쓰이는 attract, repel 개념)

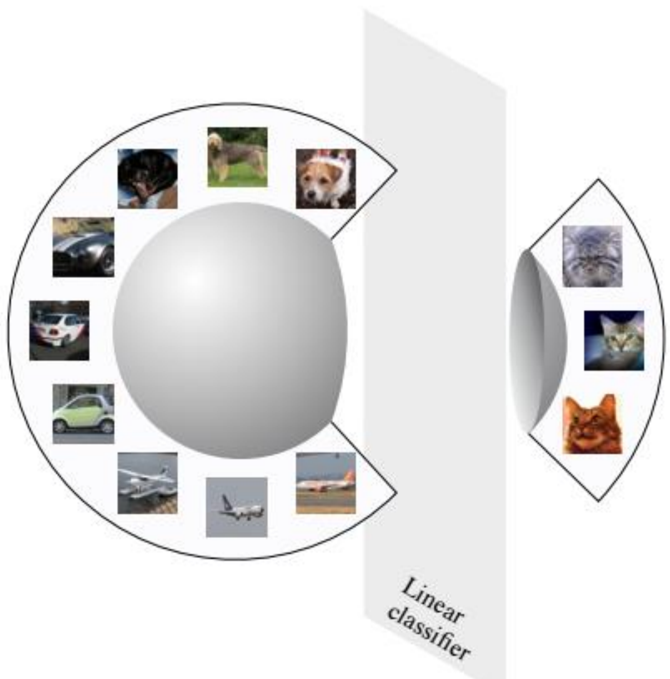
Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere

Tongzhou Wang [1] **Phillip Isola** [1]

• Introduction

❖ **Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere**

- 많은 방법론에서 l2-normalized representation을 통해 embedding 공간을 unit hypersphere 로 제한
- 이러한 방법을 통해서 학습의 안정성을 향상시키고 각 클래스 별로 linearly separable한 representation 을 학습하게 함
- 이 때, 학습되는 representation은 **불필요한 디테일에 불변**해야 하며 **가능한 많은 정보를 보존**해야 한다고 주장

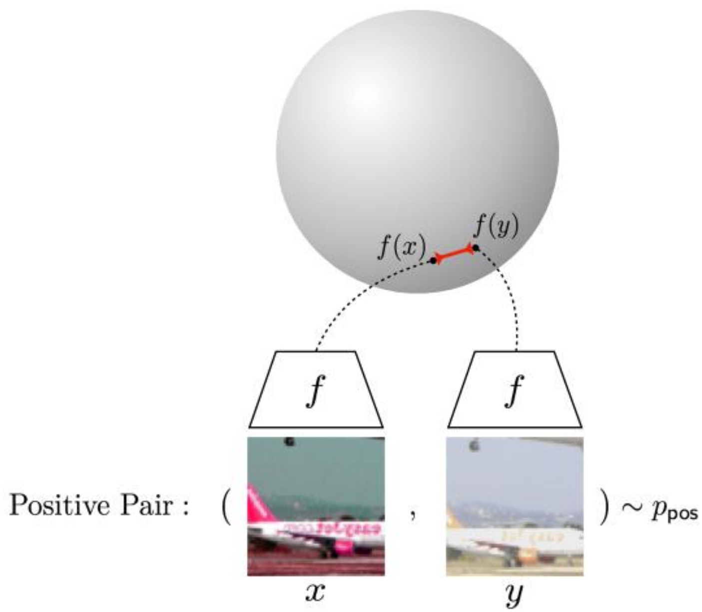


Well-clustered classes are linearly separable

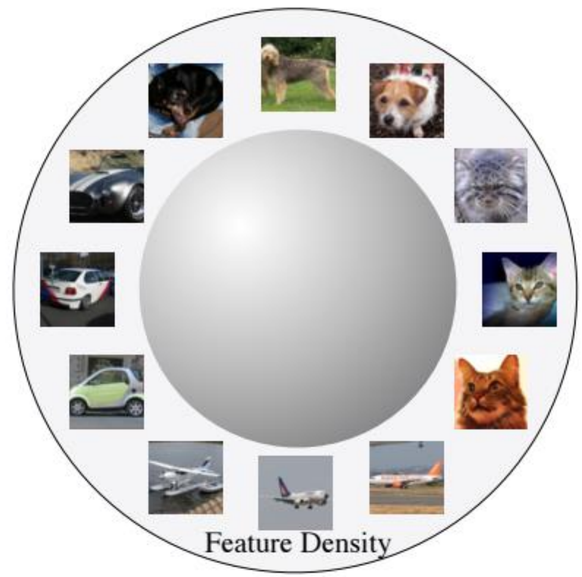
• Introduction

❖ **Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere**

- 저자들은 이러한 특성을 **Alignment**와 **Uniformity**로 정의
- **Alignment** : Unit hypersphere 상에서 positive pair 간의 가까움 정도로 비슷한 샘플은 비슷한 representation으로 mapping할수록 좋은 encoder
- **Uniformity** : Unit hypersphere 상에서 학습된 normalized representation vector 분포의 균일성으로 균일하게 분포되게 mapping할수록 좋은 encoder



Alignment



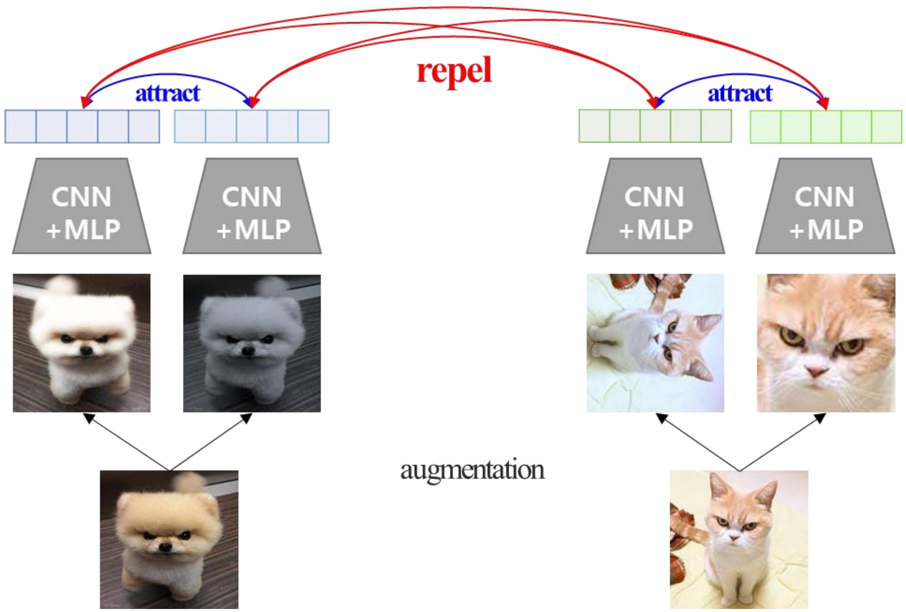
Uniformity

• Introduction

❖ Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere

- Contrastive representation learning은 augmentation을 통해 positive pair sample을 만들고 unit hypersphere 상에서 **positive pair**는 가까이 **negative samples**와는 멀도록 학습이 진행

$$\mathcal{L}_{\text{contrastive}}(f;\tau,M) \triangleq \mathop{\mathbb{E}}_{\substack{(x,y)\sim p_{\text{pos}} \\ \{x_i^-\}_{i=1}^{M} \overset{\text{i.i.d.}}{\sim} p_{\text{data}}}} \left[-\log \frac{e^{f(x)^\top f(y)/\tau}}{e^{f(x)^\top f(y)/\tau} + \sum_i e^{f(x_i^-)^\top f(y)/\tau}} \right]$$

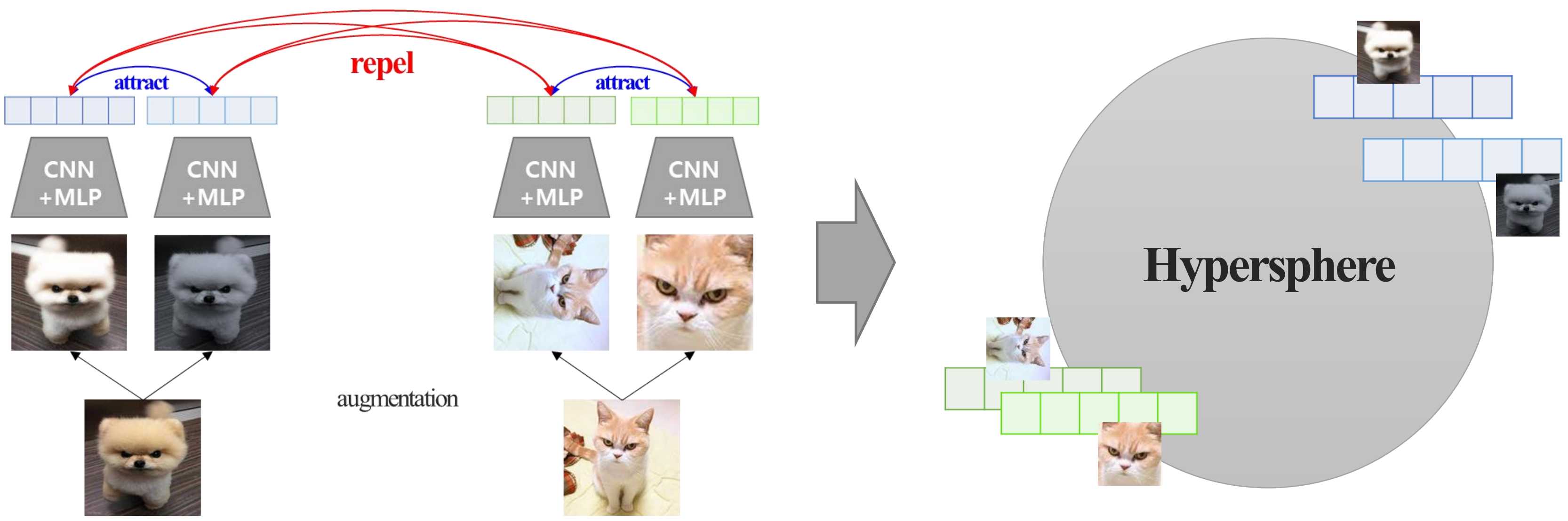


- Loss를 최소화 하기 위해서 **Positive**의 유사도는 커져야 하기 때문에 unit hypersphere에서 가까워짐 **(attract)**
- 반대로 **Negative**와의 유사도는 적어져야 하기 때문에 멀어짐 **(repel)**

• Introduction

❖ **Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere**

- 저자들은 Contrastive learning을 통해 학습되는 representation이 **alignment**와 **uniformity**를 최대화하는 방향으로 학습된다고 주장
- 이 두 특성에 대한 지표가 downstream task의 성능과 연관이 있음을 보임
- contrastive loss가 아닌 **alignment**와 **uniformity**을 직접 최적화하는 것이 더 효과적일 수 있음을 보임



• Quantifying Alignment and Uniformity

❖ Relationship between Contrastive Loss, Alignment, and Uniformity

- Contrastive loss를 변형하여 **Alignment**와 **Uniformity**의 관계를 확인
- **식 (2)**는 항상 양수 (거리 ≥ 0)이기 때문에 목적함수를 최소화 하기 위해서 **식 (1)** 이 줄어듦
- 즉, **positive pair**인 x와 y의 코사인 유사도 $f(x)^T f(y)$ 가 1에 가까워지도록 학습 (Alignment 최적화)
- **식 (1)**이 완벽하게 최적화되어 $f(x)^T f(y) = 1$이 되었다면 **식 (3)**을 최소화하도록 학습
- 즉, **Negative** (x_i^-)와의 유사도 $f(x)^T f(x_i^-)$가 0에 가까워지도록 학습 (Uniformity 최적화)

$$\mathcal{L}_{\mathsf{contrastive}}(f;\tau,M) \triangleq \mathop{\mathbb{E}}_{\substack{(x,y)\sim p_{\mathsf{pos}} \\ \{x_i^-\}_{i=1}^M \overset{\text{i.i.d.}}{\sim} p_{\mathsf{data}}}} \left[-\log \frac{e^{f(x)^\intercal f(y)/\tau}}{e^{f(x)^\intercal f(y)/\tau} + \sum_i e^{f(x_i^-)^\intercal f(y)/\tau}} \right]$$

$$= \mathop{\mathbb{E}}_{(x,y)\sim p_{\mathsf{pos}}} \left[-f(x)^\intercal f(y)/\tau \right] \quad (1)$$

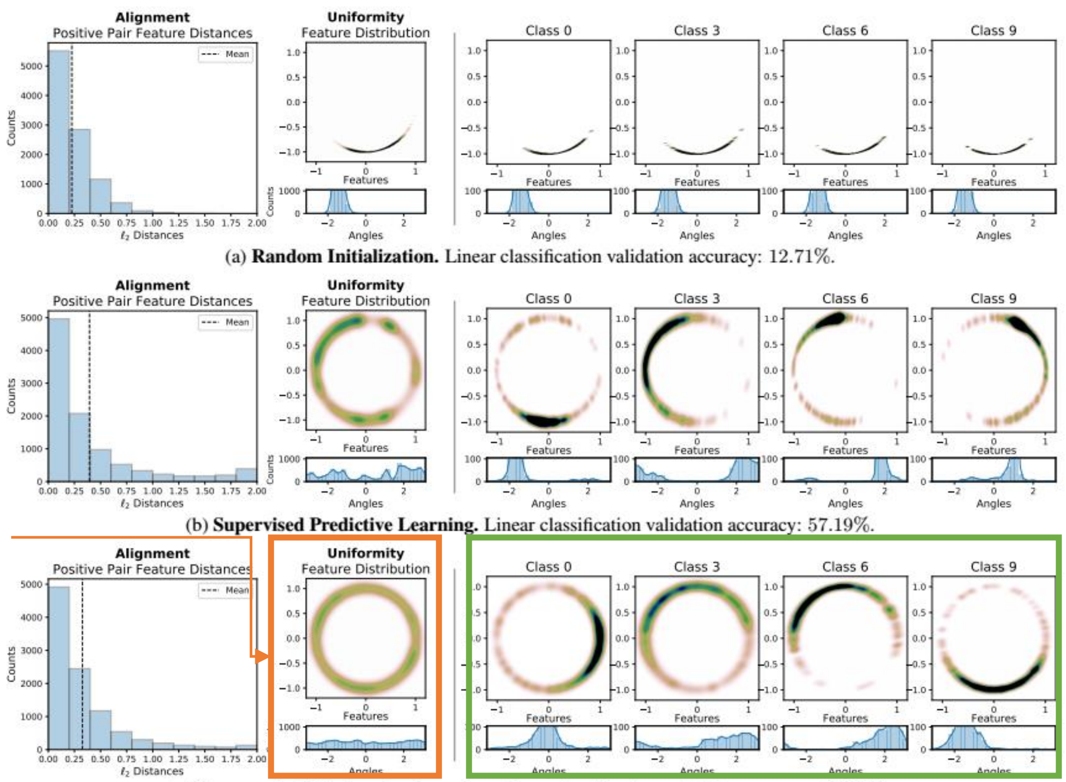
$$+ \mathop{\mathbb{E}}_{\substack{(x,y)\sim p_{\mathsf{pos}} \\ \{x_i^-\}_{i=1}^M \overset{\text{i.i.d.}}{\sim} p_{\mathsf{data}}}} \left[\log \left(e^{f(x)^\intercal f(y)/\tau} + \underbrace{\sum_i e^{f(x_i^-)^\intercal f(x)/\tau}}_{(3)} \right) \right] \quad (2)$$

DMQA

• Quantifying Alignment and Uniformity

❖ Embedding Space Visualization

- Contrastive learning으로 형성되는 embedding representation이 **alignment**와 **uniformity**를 최적화하는 방향으로 학습이 된다는 것을 증명하기 위해 CIFAR-10 데이터를 2차원으로 임베딩하여 시각화
- Contrastive learning으로 학습된 임베딩 공간에서 비슷한 샘플끼리 군집화되며 분포도 가장 균일함



(a) **Random Initialization.** Linear classification validation accuracy: 12.71%.

(b) **Supervised Predictive Learning.** Linear classification validation accuracy: 57.19%.

(c) **Unsupervised Contrastive Learning.** Linear classification validation accuracy: 28.60%.

Uniformity

- Representation의 분포가 균일함

Alignment

- 비슷한 샘플끼리 가까이 존재

DMQA

• Quantifying Alignment and Uniformity

❖ Quantification

- 더 고차원일 경우에도 contrastive learning이 **alignment**와 **uniformity**를 최적화 하도록 학습한다는 것을 증명하기 위해서 저자들은 이 두 특성을 정량화 함
- **Alignment**는 두 샘플 간 거리로 정의
- **Uniformity**는 Gaussian potential kernel (a.k.a RBF kernel)을 통해서 정의

*커널은 두 샘플의 변환된 벡터의 내적을 출력해주는 함수로 내적이기 때문에 두 벡터가 동일할 때 가장 크고 거리가 멀어질수록 값이 작아짐

$$k(x,y) = \phi(x) \cdot \phi(y)$$

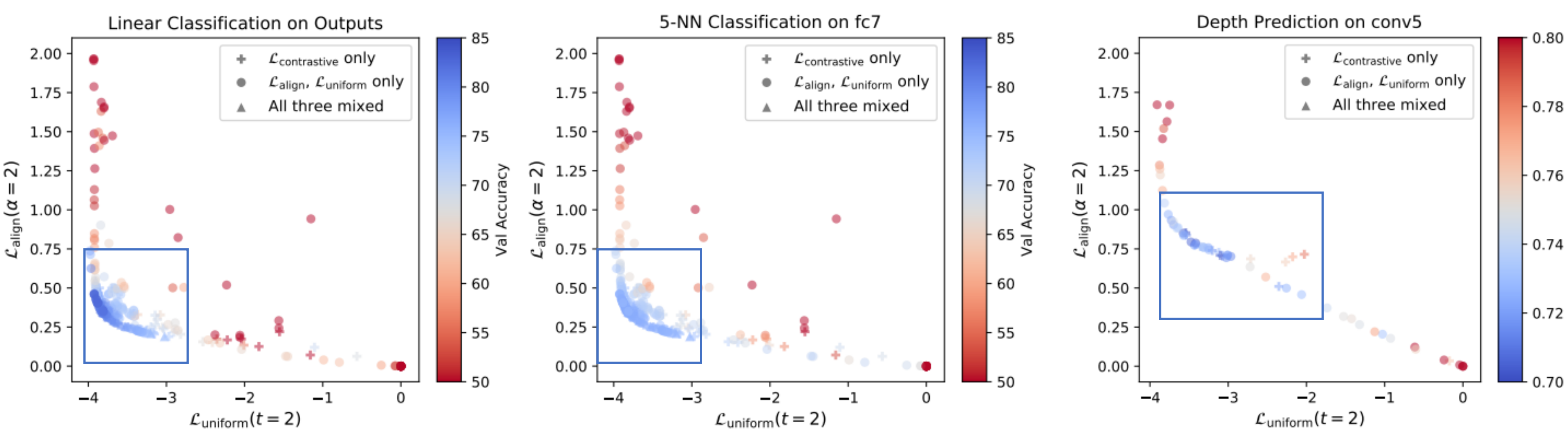
$$\mathcal{L}_{\text{align}}(f;\alpha) \triangleq \mathop{\mathbb{E}}_{(x,y)\sim p_{\text{pos}}} \left[\|f(x) - f(y)\|_2^\alpha\right], \quad \alpha > 0.$$

$$\begin{aligned}\mathcal{L}_{\text{uniform}}(f;t) &\triangleq \log \mathop{\mathbb{E}}_{x,y \overset{\text{i.i.d.}}{\sim} p_{\text{data}}} \left[G_t(u,v)\right] \\ &= \log \mathop{\mathbb{E}}_{x,y \overset{\text{i.i.d.}}{\sim} p_{\text{data}}} \left[e^{-t\|f(x)-f(y)\|_2^2}\right], \quad t > 0.\end{aligned}$$

• Experiments

❖ Relationship between Contrastive Loss, Alignment, and Uniformity

- Contrastive loss만을 사용할 때, 앞에서 정의한 Alignment와 uniformity 함수를 loss로 사용할 때, 세 함수 모두를 목적함수로 사용할 때의 alignment와 uniformity 값을 비교
- Alignment(y축)와 uniformity(x축)의 값이 모두 작을 (왼쪽 모서리) 경우 가장 성능이 좋으며(파란색), contrastive loss만을 쓸 때에도 두 값이 작을 때 성능이 좋음



DMQA

• Experiments

❖ **Relationship between Contrastive Loss, Alignment, and Uniformity**

- 또한, contrastive loss를 사용하지 않고 본 논문에서 정의한 alignment와 uniformity를 직접 목적함수로 사용하여 최적화하는 것이 약간의 성능향상을 보임

	Loss Formula	Validation Set Accuracy ↑			
		Output + Linear	Output + 5-NN	fc7 + Linear	fc7 + 5-NN
Best $\mathcal{L}_{\mathsf{contrastive}}$ only	$\mathcal{L}_{\mathsf{contrastive}}(\tau{=}0.19)$	80.46%	78.75%	83.89%	76.33%
Best $\mathcal{L}_{\mathsf{align}}$ and $\mathcal{L}_{\mathsf{uniform}}$ only	$0.98 \cdot \mathcal{L}_{\mathsf{align}}(\alpha{=}2) + 0.96 \cdot \mathcal{L}_{\mathsf{uniform}}(t{=}2)$	**81.15%**	78.89%	**84.43%**	**76.78%**
Best among all encoders	$\mathcal{L}_{\mathsf{contrastive}}(\tau{=}0.5) + \mathcal{L}_{\mathsf{uniform}}(t{=}2)$	81.06%	**79.05%**	84.14%	76.48%

Table 1: STL-10 encoder evaluations. Numbers show linear and 5-nearest neighbor (5-NN) classification accuracies on the validation set. The best result is picked by encoder outputs linear classifier accuracy from a 5-fold training set cross validation, among all 150 encoders trained from scratch with 128-dimensional output and 768 batch size.

	Loss Formula	Validation Set MSE ↓	
		conv5	conv4
Best $\mathcal{L}_{\mathsf{contrastive}}$ only	$0.5 \cdot \mathcal{L}_{\mathsf{contrastive}}(\tau{=}0.1)$	0.7024	**0.7575**
Best $\mathcal{L}_{\mathsf{align}}$ and $\mathcal{L}_{\mathsf{uniform}}$ only	$0.75 \cdot \mathcal{L}_{\mathsf{align}}(\alpha{=}2) + 0.5 \cdot \mathcal{L}_{\mathsf{uniform}}(t{=}2)$	**0.7014**	0.7592
Best among all encoders	$0.75 \cdot \mathcal{L}_{\mathsf{align}}(\alpha{=}2) + 0.5 \cdot \mathcal{L}_{\mathsf{uniform}}(t{=}2)$	**0.7014**	0.7592

Table 2: NYU-DEPTH-V2 encoder evaluations. Numbers show depth prediction mean squared error (MSE) on the validation set. The best result is picked based on conv5 layer MSE from a 5-fold training set cross validation, among all 64 encoders trained from scratch with 128-dimensional output and 128 batch size.

• Experiments

❖ Relationship between Alignment and Uniformity

- Alignment와 Uniformity의 각 목적함수 별 weight를 변경해 가면서 실험
- 맨 왼쪽은 alignment 목적함수만 사용한 것이며 맨 오른쪽은 uniformity 목적함수만 사용
- 양 극단으로 갈수록(하나의 목적함수만 사용) 성능이 급격히 떨어지며 두 목적함수를 적절하게 같이 사용하는 것이 좋은 성능을 보임
- 이를 통해 alignment와 uniformity가 동시에 최적화 되어야 좋은 성능을 보임을 확인

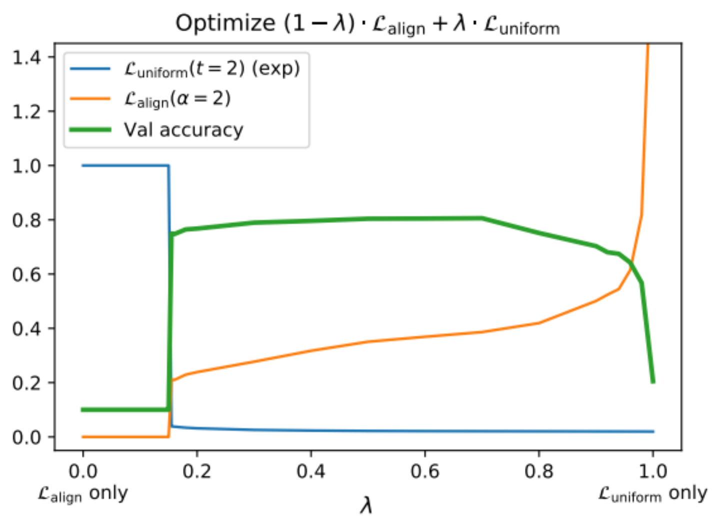


Figure 7: Effect of optimizing different weighted combinations of $\mathcal{L}_{\mathrm{align}}(\alpha=2)$ and $\mathcal{L}_{\mathrm{uniform}}(t=2)$ for STL-10. For each encoder, we show the $\mathcal{L}_{\mathrm{align}}$ and $\mathcal{L}_{\mathrm{uniform}}$ metrics, and validation accuracy of a linear classifier trained on encoder outputs. $\mathcal{L}_{\mathrm{uniform}}$ is exponentiated for plotting purposes.

DMQA

• Summary & Conclusions

❖ Summary

- 본 논문에서는 contrastive representation learning이 alignment와 uniformity를 최적화하는 방향으로 학습된다고 주장
- 이 증명을 위해 alignment와 uniformity를 l-norm과 가우시안 필터로 정량화한 뒤 다양한 실험을 통해서 자신의 주장을 증명
- 이를 통해 기존에 Mutual Information을 최대화한다는 해석보다 더욱 직관적으로 해석하고 이해하는 것에 도움을 줌

❖ Conclusions

- Contrastive learning과 alignment, uniformity 간의 관계를 보여주는 수식 전개 과정은 너무 어려워서 넘어갔음
- 하지만 Representation learning이 어떻게 학습되는지 시각적으로 보여줘서 논문 제목 그대로 이해하기 좋았음
- 다만 두 목적함수를 사용한다면 weight search가 힘들 것 같아서 그냥 contrastive loss를 사용하는 것이 더 좋아 보였음

• Reference

❖ Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere

1. Wang, Tongzhou, and Phillip Isola. "Understanding contrastive representation learning through alignment and uniformity on the hypersphere." *International Conference on Machine Learning*. PMLR, 2020.
2. https://velog.io/@kimkj38/%EB%85%BC%EB%AC%B8-%EB%A6%AC%EB%B7%B0-Understanding-Contrastive-Representation-Learning-through-Alignment-and-Uniformity-on-the-Hypersphere

DMQA

• Appendix A

❖ **proof of the positive semi-definite of Laplacian**

임의의 $x \in \mathbb{R}^N$ 에 대해,

$$\begin{aligned}
x^T L x = x^T D x - x^T W x &= \sum_i D_{ii} x_i^2 - \sum_{i,j} x_i W_{ij} x_j \\
&= \frac{1}{2}\left(2\sum_i D_{ii} x_i^2 - 2\sum_{i,j} W_{ij} x_i x_j\right) \\
&\overset{(a)}{=} \frac{1}{2}\left(2\sum_i \left\{\sum_j W_{ij}\right\} x_i^2 - 2\sum_{i,j} W_{ij} x_i x_j\right) \\
&\overset{(b)}{=} \frac{1}{2}\left(\sum_{i,j} W_{ij} x_i^2 + \sum_{i,j} W_{ij} x_j^2 - 2\sum_{i,j} W_{ij} x_i x_j\right) \\
&= \frac{1}{2}\sum_{i,j} W_{ij}(x_i - x_j)^2 \geq 0 \qquad (2)
\end{aligned}$$

https://harryjo97.github.io/theory/Graph-Laplacian/

DMQA

Thank You