각도 마진 손실 함수를 적용한 객체 분류

박선지 *조남익
서울대학교 전기정보공학부, 뉴미디어통신공동연구소
seonjipark@ispl.snu.ac.kr, *nicho@snu.ac.kr

Object Classification with Angular Margin Loss Function

Seonji Park, *Namik Cho
Department of ECE, INMC, Seoul National University

요 약

객체 분류는 입력으로 주어진 이미지에 포함된 객체의 종류를 판단하는 기술이다. 대표적인 딥러닝 기반의 객체 분류 방법으로서 Faster R-CNN[2], YOLO[3] 등의 모델이 개발되었으나, 여전히 성능 향상의 여지가 있다. 본 연구에서는 각도 마진 손실 함수를 기존의 몇 가지 객채 분류 모델에 적용하여 성능 향상을 유도한다. 각도 마진 손실 함수는 얼굴 인식 모델인 SphereFace [4]에서 제안한 방법으로, 얼굴 인식과 같이 단일 도메인의 데이터셋을 분류하는 문제를 풀기 위해 제안되었다. 이는 기존 소프트맥스 함수에서 클래스 결정 경계선에 마진을 주는 방식으로 클래스 간의 구분 능력을 향상시킨다. 본 논문은 각도 마진 손실 함수를 CIFAR10, CIFAR100 데이터셋의 분류 문제에 적용하였으며 ResNet, EfficientNet, MobileNet 등의 백본 네트워크로 실험하여 평균적으로 mAP 성능이 향상되는 것을 확인하였다.

1. 서론

객체 분류는 입력으로 주어진 이미지에 포함되어 있는 객체가 어떤 종류에 속하는 지 판단하는 기술로서, 객체 인식, 객체 의미 분할 등과 함께 대표적인 이미지 분석 기술 중 하나이다. 2015 년도에 객체 인식 및 분류 대회인 ILSVRC 에서 ResNet[1]이 사람의 객체 분류 정확도 오차범위인 5%보다 더 낮은 오차율을 달성하면서 딥러닝 기반의 객체 분류 기술이 활발하게 연구되어 왔다. 같은 해에 Faster R-CNN[2]과 YOLO[3]가 등장하면서 그 후속 연구들이 객체 분류 연구에서 주류를 이루고 있다. 특히 [2], [3]은 분류성능과 함께 실용성을 위한 빠른 분류 속도를 목표로 하여 이들 모두 분류 속도를 약 45 FPS 로 개선하였으나 분류 정확도 측면에서는 여전히 개선의 여지가 있다. 이에 따라 ResNet[1] 등의 기본적인 컨볼루션 네트워크를 기반으로 하여 피쳐 피라미드 네트워크(FPN)와 같이 구조를 변형하거나, 마진 손실 함수와 같이 손실 함수를 개선하는 연구도 진행되고 있다.

손실 함수를 개선하는 연구 중 각도 마진 손실 함수는 일반적인 객체 분류 문제가 아닌 얼굴 인식 문제에서 먼저 등장하였다. SphereFace[4]가 최초로 기존의 소프트맥스 함수를 변형한 후 마진을 추가하여 분류 성능을 높인 각도-소프트맥스 함수를 제안하였다. CosFace[5]와 ArcFace[6]가 뒤이어 각각 SphereFace[4]와는 다른 방식으로 마진을 추가하는 방법을 제안하였다. 각도 마진 손실 함수(Angular margin loss function)에 대한 내용은 2 절에서 더욱 자세히 다루도록 한다.

본 논문에서는 기존 얼굴 인식 기술에서 사용되던 각도 마진 손실 함수를 일반적인 객체 분류 문제에 적용하여 기존의 선형 레이어를 통과한 후 크로스 엔트로피 손실 함수를 사용하는 경우보다 성능 향상을 이루었다.

본 논문의 구성은 다음과 같다. 2 절에서는 각도 마진 손실 함수에 대해 살펴본 후, 3 절에서는 본 논문에서 제안하는 각도 마진 손실 함수를 일반적인 객체 분류 데이터셋인 CIFAR10 과 CIFAR100 에 적용하는 실험 방법에 대해 설명한다. 4 절에서는 실험 결과에 대해 정리하고, 5 절에서는 본 논문에 대한 결론을

맺는다.

2. 각도 마진 손실 함수

각도 마진 손실 함수는 얼굴 인식 모델 중 하나인 SphereFace[4]에서 제안한 기술로, 기존 소프트 맥스 손실 함수에 입력으로 들어가는 모델 weight 를 정규화한 후, 모델의 출력물인 logit 에 추가적인 마진 가중치를 곱해 그 분류 능력을 향상시키는 방법이다. 얼굴 인식 모델은 다양한 도메인의 객체를 구분하는 일반적인 객체 분류 모델과 달리 얼굴이라는 하나의 도메인에 속한 데이터셋에서 입력 얼굴 이미지가 어떤 사람의 얼굴인지를 분류해야 한다. 또한 분류 대상이 되는 얼굴의 종류가 수천 개 단위로 상당히 많다. 따라서 일반 객체 분류보다 더욱 세밀하고 보수적인 분류 기준이 필요하다.

이러한 목적을 달성하기 위해 SphereFace[4]는 먼저 기존의 소프트맥스 함수에 입력으로 들어가는 모델 weight 와 입력 피쳐 x 에 L2 정규화를 하여 결정 경계선을 마치 2 차원 상의 각좌표계처럼 해석한다. 먼저 기존 소프트맥스 함수의 식은 아래와 같다.

$$L = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{W_{y_j}^T x_i + b_j}}{\sum_{j=i}^{n} e^{W_j^T x_i + b_j}}, \quad (1)$$

이때, x_i는 y_i 클래스에 속한 i 번째 입력 이미지의 피쳐이며 W_j는 모델 weight W 의 j 번째 열, b_j는 bias 를 의미한다. N 은 입력 배치의 사이즈이며 n은 입력 데이터셋의 클래스 개수이다.

식 (1) 을 단순화 하기 위해 먼저 $\|W\|$ =1, $\|x\|$ = s(s는 임의의 실수) 으로 각각 정규화하면 weight W 와 입력 x 사이의 내적 식은 아래와 같이 정리할 수 있다.

$$W_{y_j}^T x_i = \|W_j\|\|x_i\| cos\theta_j \cong s * cos\theta_j, \quad (2)$$

식 (2)와 같이 소프트맥스 손실 함수에 정규화한 W, x를 사용하여 표현하는 것을 정규화-소프트맥스 손실함수[4]라고 부르며 이때 최종 손실 함수는 아래와 같다.

$$L = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{s*cos\theta_{y_i}}}{e^{s*cos\theta_{y_i}} + \sum_{j=i,j\neq y_i}^{n} e^{s*cos\theta_j}}, \quad (3)$$

SphereFace[4]에서 추가적으로 제안하는 방법은 각도 θ_j 에 직접 마진 m (m>0) 을 곱하여 이미지 피쳐의 결정 경계선에 마진을 주어 그 분별력을 높이는 것으로서 아래와 같은 식으로 정리된다.

$$L = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{cos(m\theta_{y_i})}}{e^{cos(m\theta_{y_i})} + \sum_{j=i,j\neq y_i}^{n} e^{cos\theta_j}}, \quad (4)$$

기존 소프트 맥스 함수는 결정 경계선을 기준으로 클래스를 분류했지만, 각도 마진 손실 함수를 사용하면 그림 1 에서 회색 영역으로 표시된 부분과 같이 마진에 해당하는 범위 이상으로 feature representation 에 차이가 있어야 해당 클래스로 분류될 수 있다. 예를 들어 클래스 1, 2 만 존재하는 데이터셋에서 하나의 데이터가 클래스 1 로 분류되기 위해서는 클래스 2 와의 feature representation 차이가 그림 1 에서 보는 바와 같이 마진 m 보다 커야 한다.

이때, ArcFace[6]은 SphereFace[4]가 제안하는 마진 부과 방식으로 최적화된 피쳐를 기하학적으로 분석하면 평행한 선형 결정 경계선을 갖지 않는다는 점을 문제로 지목했다(그림 1). 각도에 마진을 곱해 경계선에 마진을 부과하면 클래스별 피쳐 클러스터의 중심각에 마진을 준 것 일뿐 실제 경계선 사이의 직교 거리가 멀어지는 것이 아니기 때문이다. 이에 따라 ArcFace[6]는 클래스 분류 경계선에 평행하게 마진을 주기 위해서 식 (3)에서 추가적인 마진을 각도에 더해주는 방식을 제안한다. 이때의 손실 함수 식은 아래 식 (5)와 같다.

$$L = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{cos(\theta_{y_i} + m)}}{e^{cos(\theta_{y_i} + m)} + \sum_{j=i,j\neq y_i}^{n} e^{cos\theta_j}}, \quad (5)$$

각 방법이 지향하는 최적화된 피쳐를 기하학적으로 표현하면 아래 그림 1 과 같다. 그림은 단순한 예시로 분류 대상이 되는 클래스가 2 종류(클래스 1, 클래스 2)인 경우의 경계선을 표현한 것이다.

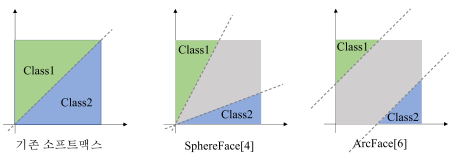

그림 1. 소프트맥스와 각도 마진 손실 함수[4,6]의 feature representation. 회색 영역이 각 클래스 사이의 마진 영역을 나타낸다.

최종적으로 본 논문에서는 ArcFace[6]의 각도 마진 손실 함수를

사용한다.

3. 실험 방법

본 논문은 분류 데이터셋으로 CIFAR10, CIFAR100[9]을 사용했다. CIFAR10, CIFAR100 은 각각 10 개, 100 개의 클래스에 대한 이미지를 포함하고 있으며, 클래스의 예시로는 비행기, 강아지, 말, 트럭 등이 있다. CIFAR100 은 사람, 꽃, 곤충의 세세한 종류를 나타내는 더욱 다양한 클래스의 데이터를 포함하고 있다.

분류 백본 네트워크로는 ResNet[1]과 함께 대표적인 경량화 컨볼루션 네트워크인 MobileNet[7], EfficientNet[8]을 베이스라인으로 사용했다. CIFAR10, CIFAR100[9] 데이터셋이 일반적으로 객체 분류 데이터셋 중에서 소규모 데이터셋에 해당하는 점을 고려해 적은 수의 파라미터로도 충분히 성능을 낼 수 있는 경량화 네트워크[7], [8]를 사용하였다.

각도 마진 손실 함수는 입력 x 의 정규화 값을 조절하는 파라미터 s 와 마진의 정도를 조절하는 파라미터 m 을 하이퍼파라미터로 갖는다. 다양한 s, m 값을 이용하여 실험한 결과 s = 1, m = 0.1 을 사용하였을 때 가장 좋은 결과를 얻을 수 있었다.

4. 실험 결과

객체 분류 성능은 mAP 를 이용하여 측정하였다. 일부 실험에서는 기존의 방법보다 약간 떨어지는 성능을 보였지만, 대체로 각도 마진 손실 함수를 사용한 경우에 성능 향상을 확인할 수 있었다.

표 1 CIFAR10

	ResNet18	EfficientNet	MobileNetV2
크로스엔트로피	65.2	84.66	**59.01**
각도마진손실함수	**65.4**	**85.2**	58.88

표 2 CIFAR100

	ResNet[1]	EfficientNet[8]	MobileNet[7]
크로스엔트로피	**60.16**	62.2	58.94
각도마진손실함수	59.8	**63.01**	**59.05**

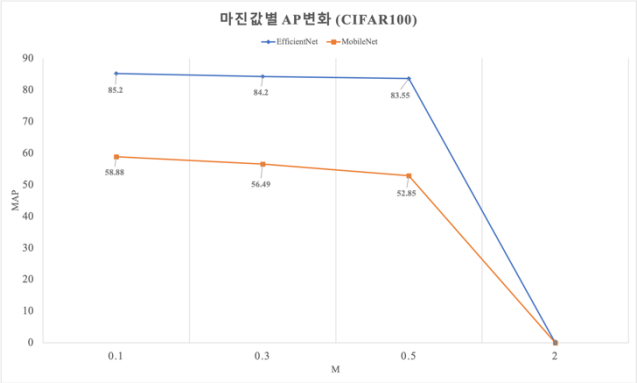

그림 2. 마진 m 값 변화시 mAP 변화 추이

그림 2 는 최적의 마진 m 값을 찾기 위해 CIFAR10 데이터셋에서 m 값을 변형시켜가며 학습시켰을 때의 mAP 값 변화를 표현한 그림이다. m=2 이상의 값을 사용했을 때에는 허용 가능한 마진 범위를 넘어서 학습이 수렴하지 않거나 테스트셋의 성능이 나오지 않는 것을 확인할 수 있었다. EfficientNet, MobileNet 모두 m = 0.1 에서 가장 좋은 mAP 값을 보였다.

CIFAR10, CIFAR100[9] 이외에도 대규모 데이터셋 중 하나인 ImageNet 을 이용하여서도 학습을 진행해보았다. 다만 1 천개 이상의 클래스를 가진 ImageNet 데이터셋에서는 ResNet50 등의 큰 백본 네트워크를 사용하여도 학습이 수렴하지 않는 것을 확인할 수 있었다. 큰 데이터셋을 사용하는 경우에는 기존의 크로스엔트로피 손실 함수와 각도 마진 손실 함수를 함께 사용하는 등의 보완하는 연구를 차후에 진행할 예정이다.

5. 결론

본 논문에서는 객체 분류 모델의 분류 정확도를 높이기 위해 각도 마진 손실 함수를 사용할 것을 제안한다. 각도 마진 손실 함수는 의사 결정 경계선만을 이용하던 기존의 소프트맥스 함수에 비해 더욱 정확하게 클래스를 분류하는 것을 목표로 하며, 실제로도 CIFAR10, CIFAR100 에 이를 적용하여 기존 백본 네트워크보다 mAP 성능이 향상되는 것을 확인하였다. 다만 1 천개 이상의 클래스를 가진 대규모 데이터셋을 사용할 경우 학습이 불안정하다는 한계를 갖고 있어서 추후 연구를 통해 보완이 필요해 보인다.

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 정보통신·방송 연구개발사업의 일환으로 수행하였으며 2022년도 BK21 FOUR 정보기술 미래인재 교육연구단에 의하여 지원되었음. [2021-0-01062-0 01 , 자율주행용 수집/활용 데이터에 대한 개인정보 처리 기술개발]

참 고 문 헌

[1] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
[2] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems* 28 (2015).
[3] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
[4] Liu, Weiyang, et al. "Sphereface: Deep hypersphere embedding for face recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
[5] Wang, Hao, et al. "Cosface: Large margin cosine loss for deep face recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
[6] Deng, Jiankang, et al. "Arcface: Additive angular margin loss for deep face recognition." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.
[7] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
[8] T an, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." *International conference on machine learning*. PMLR, 2019.
[9] Krizhevsky, Alex, and Geoffrey Hinton. "Learning multiple layers of features from tiny images." (2009): 7.