

Feature Similarity Contrastive Learning 기반 불균형 Fine-Grained 야생동물 이미지 분류 방법

이서희¹, 박가람¹, 손규원¹, 도희찬¹, 양희성², 박혜영²

¹ 경북대학교 전자공학부

² 경북대학교 컴퓨터학부

{ds3421, rkfka2430, dreamsuga1, jsinvade392, hs.yang, hypark}@knu.ac.kr

Feature Similarity Contrastive Learning for Imbalanced Fine-Grained Wildlife Image Classification

Seohee Lee¹, Garam Park¹, Gyuwon Son¹, Heechan Do¹,

Heesung Yang², Hyeyoung Park²

¹School of Electronics Engineering, Kyungpook National University

²School of Computer Science and Engineering, Kyungpook National University

요 약

야생동물 영상의 종 분류는 생태계 보전 정책 수립에 필수적이지만, 데이터 불균형과 fine-grained 클래스 간 유사성으로 인해 높은 정확도를 달성하기 어렵다. 본 연구에서는 데이터 불균형 문제와 fine-grained 클래스 분류의 어려움을 동시에 해결하기 위한 새로운 학습 방법을 제안한다. 제안하는 방법은 oversampling과 중간 계층 feature 유사도에 기반한 contrastive loss를 기존의 영상 분류 신경망 모델과 결합하여, 분류 학습을 수행하는 동시에 클래스 내 feature의 유사성과 클래스 간 feature의 비유사성을 동시에 학습한다. 제안하는 방법을 실제 야생동물 영상 데이터셋을 대상으로 검증한 결과, coarse-grained와 fine-grained 조건 모두에서 기존의 방법보다 높은 클래스 분류 정확도를 달성하였다. 특히, 심각한 데이터 불균형 조건에서도 안정적인 성능을 유지하였으며, 효과적인 feature 분리와 함께 별도의 fine-tuning 없이 end-to-end 학습이 가능함을 실험적으로 입증하였다. 본 연구에서 제안하는 방법은 향후 생태계 모니터링 뿐 아니라 다양한 분야의 fine-grained 희귀 클래스 인식 분야에도 적용할 수 있을 것으로 기대된다.

1. 서 론

야생동물 영상의 종 분류는 생태계 구조를 이해하고 보전 정책을 수립하는 데 필수적이다. 그러나 이 작업은 두 가지 주요 문제로 인해 어려움을 겪는다. 첫째, 자연 환경에서 수집된 영상 데이터는 종 간 출현 빈도의 심각한 불균형(data imbalance)을 포함하며, 이는 심층 학습 기반 분류기의 성능 저하를 초래한다 [1]. 둘째, 국내 야생동물 데이터셋은 외형적으로 유사한 종(fine-grained class)을 포함하고 있어 정확한 분류를 어렵게 한다. 예를 들어, 노루(roe deer)와 고라니(water deer)는 계통학적으로도 가까워 유사한 크기와 형태를 지니기 때문에 식별이 어렵다 [2]. 따라서 야생동물 영상 분류 문제는 데이터 불균형과 외형적으로 유사한 종 간의 구분이라는 두 가지 복합적인 문제로 정의될 수 있다.

각 문제를 해결하기 위해 다양한 접근법이 시도되어 왔다. 먼저 클래스 간 샘플 수의 불균형을 완화하는 방법으로는 학습 중 샘플 가중치를 적용하거나 무작위 데이터 샘플링을 수행하는 oversampling 기법이 사용된다 [3]. 또한, 손실 함수 설계

측면에서는 클래스별 유효 샘플 수를 고려하여 소수 클래스에 대한 보정을 수행하는 Class-Balanced Loss가 대표적이나, 이 방식은 oversampling 기법과 병행 사용이 어렵다는 한계를 가진다 [4]. 외형적으로 유사한 종을 분류하기 위해서는 클래스 내 유사성과 클래스 간 비유사성을 동시에 학습하는 supervised contrastive learning 방식이 활용된다. 이 방법은 외형적으로 유사한 클래스 간의 특징을 효과적으로 추출할 수 있으며, 동시에 데이터 불균형 문제에도 효과적으로 대응한다 [5]. 하지만, 이러한 contrastive learning 방법은 학습 이후 별도의 fine-tuning 단계를 요구하여 end-to-end 학습이 불가능하다.

본 연구는 이러한 기존 방법들의 한계를 극복하며, 두 가지 복합적인 문제를 동시에 효과적으로 해결할 수 있는 새로운 학습 전략을 제안한다. 제안하는 방법은 fine-grained 클래스 간 차별적 특징 추출을 촉진하기 위해, 중간 계층 feature의 유사도에 기반한 contrastive loss를 기존의 분류 손실 함수에 추가한다. 이 접근법은 다양한 데이터 샘플링 기법과 함께 사용할 수 있어 다양한 불균형 상황에 적용 가능하며, 기존 contrastive learning 기반 방법들과 달리 별도의 fine-tuning 없이 end-to-end 학습을 가능하게 한다. 본 연구는 이러한 기존 방법들의 한계를 극복하며, 두 가지 본 연구에서는 제안하는

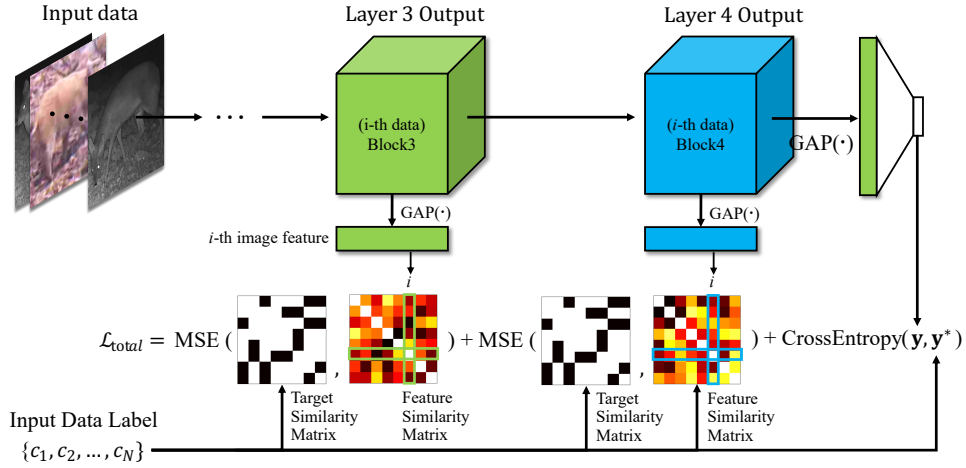


그림 1 제안하는 학습 방법의 구조

방법의 효과를 실제 야생동물 이미지 데이터셋 [6]에 다양한 데이터 불균형 조건에서 검증하였다.

2. 제안하는 방법

본 연구에서는 야생동물 영상에서 발생하는 fine-grained 클래스 간 차별적인 특징을 효과적으로 학습하기 위해 contrastive loss를 cross-entropy loss와 결합하는 학습 방법을 제안한다. 추가적으로, 제안하는 방법은 불균형 데이터에 대하여 oversampling을 적용하기 위해 여러 데이터 증강을 함께 사용한다.

그림 1에 제안하는 학습 구조를 나타낸다. oversampling을 통해 얻어진 이미지 배치는 backbone 신경망의 입력으로 주어지며, 학습 과정에서 중간 층에서 feature plane들이 추출된다. 각 중간 층에서 추출된 feature plane은 global average pooling (GAP)을 거쳐 배치내의 각 이미지에 대한 feature vector로 변환된다. 신경망의 l 층에서 얻어진 i 번째 이미지의 feature vector \mathbf{h}_i^l 에 대하여 배치 내 특징 유사도 행렬 \mathbf{S}^l 은 식 (1)을 통해 계산할 수 있다.

$$\mathbf{S}^l = [s_{ij}^l] \\ s_{ij}^l = \frac{\mathbf{h}_i^l \cdot \mathbf{h}_j^l}{\|\mathbf{h}_i^l\| \|\mathbf{h}_j^l\|} \quad (1)$$

배치 내 i 번째 이미지의 클래스 라벨 정보 c_i 에 대하여 배치 내 타겟 특징 유사도 행렬 \mathbf{T} 는 식 (2)를 통해 계산한다.

$$\mathbf{T} = [t_{ij}] \\ t_{ij} = \begin{cases} 1 & \text{if } c_i = c_j \\ 0 & \text{if } c_i \neq c_j \end{cases} \quad (2)$$

식 (2)에서 c_i 와 c_j 가 동일하지 않을 때 -1이 아닌 0을 사용하는 이유는 다른 클래스의 이미지라도 유사한 특징을 가지는 것을 허용하기 위함이다. 예를 들어, s_{ij}^l 가 -1이 된다면, i 번째 이미지와 j 번째 이미지는 추출된 특징에 대하여 반대 방향으로 존재하며, 이는 분류 성능을 하락 시키는 요인이 될 수 있다. 이후 식 (1)과 (2)에서 얻어진 유사도 행렬들과 신경망의 출력 \mathbf{y} , 타겟 벡터 \mathbf{y}^* 를 이용하여 식 (3)과 같이 l 층의 contrastive loss

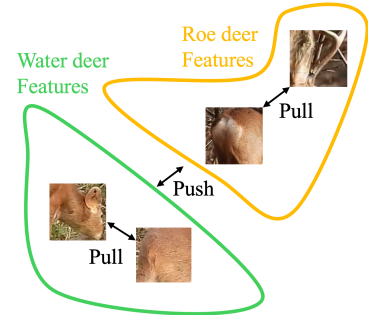


그림 2 제안하는 방법의 특징 공간 구조

\mathcal{L}_c^l 과 전체 loss \mathcal{L} 를 계산할 수 있다.

$$\mathcal{L}_c^l = \text{MSE}(\mathbf{S}^l, \mathbf{T}) \\ \mathcal{L}_{\text{total}} = \text{CrossEntropy}(\mathbf{y}, \mathbf{y}^*) + \sum_l \mathcal{L}_c^l \quad (3)$$

그림 2는 feature 공간에서 제안하는 방법의 작동 방식을 나타낸다. 제안하는 방법은 fine-grained 클래스의 차별적인 특징을 추출하기 위해, 동일 클래스의 feature는 유사하게, 상이한 클래스의 feature는 유사하지 않도록 조정한다. 이러한 방식은 추출된 feature가 fine-grained 클래스 간의 공통된 특징이 아닌, 판별에 유리한 차별적인 특징을 다음 층으로 전달하도록 한다.

3. 실험 결과

제안하는 방법의 성능을 검증하기 위해 야생동물 활동 영상 데이터 [6]를 사용하였다. 실험은 두 가지 조건에서 수행하였다. 첫째, 고라니(water deer)와 멧돼지(wild boar)를 1:1, 1:0.1, 1:0.01, 1:0.005, 1:0.001의 비율로 조정하여 imbalanced 조건에서 성능을 평가하였다. 둘째, 고라니와 노루(roe deer)를 동일한 비율로 조정하여 imbalanced fine-grained 조건에서 검증하였다. 두 실험 모두 ResNet-18 [7]을 기본 모델로 사용하였으며, 배치사이즈는 64, 학습률은 0.001로 설정하고 100회 동안 학습하였다. 학습 데이터에서 고라니 영상은 두 조건 모두 10,000개로 고정하였으며, 테스트 데이터는 고라니

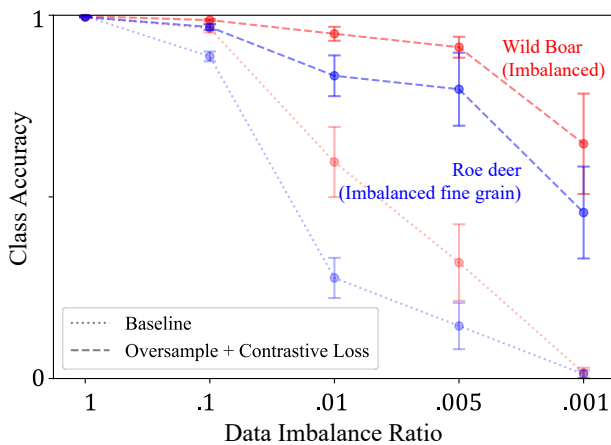


그림 3 데이터 불균형 비율에 따른 클래스별 정확도

표 1 불균형 데이터 조건에서 방법 별 클래스 정확도

Accuracy for wild boar class (Coarse-grained)					
Data ratio	1	0.1	0.01	0.005	0.001
Baseline	0.9937	0.9632	0.5988	0.3704	0.0058
Contrastive	0.9981	0.9759	0.7237	0.4774	0
Oversampling	0.9977	0.9907	0.9247	0.9643	0.5953
Proposed	0.9981	0.9898	0.9319	0.9444	0.5133

Accuracy for roe deer class (Fine-grained)					
Data ratio	1	0.1	0.01	0.005	0.001
Baseline	0.9978	0.8721	0.3409	0.155	0.0067
Contrastive	0.9961	0.9281	0	0	0
Oversampling	0.9948	0.9627	0.9393	0.7771	0.4348
Proposed	0.9944	0.9604	0.8859	0.907	0.3117

4,119개, 멧돼지 4,317개, 노루 3,142개 영상으로 구성하였다. 학습 과정에서 데이터 불균형을 완화하기 위해 소수 클래스 데이터를 다수 클래스와 1:1 비율로 oversampling 하였고, 추가적으로 RandAugment [8]와 color jittering을 적용하여 데이터 다양성을 확보하였다. 마지막으로, 성능 평가는 불균형 클래스에 대한 분류 정확도를 기준으로 평가하였다.

그림 3에 데이터 불균형 비율에 따른 클래스별 평균 정확도와 표준 편차를 나타내었다. 각 실험은 10회 진행되었으며, baseline 방법은 cross-entropy loss만을 사용한 모델이며, 제안하는 방법은 oversampling과 contrastive loss를 함께 적용하였다. 실험 결과는 fine-grained 클래스인 roe deer와 coarse 클래스인 wild boar 모두에서, 데이터 불균형 비율이 심화될수록 baseline 방법의 정확도는 급격히 감소하였다. 특히, fine-grained 클래스의 경우 baseline은 데이터 불균형 비율이 0.01 이하로 낮아질 때 정확도가 크게 하락하였다. 반면, 제안하는 방법은 모든 불균형 조건에서 baseline 대비 높은 정확도를 유지하였다. 특히, 10,000개의 데이터에서 50개의 데이터(0.005 ratio)까지 모든 경우에서 일반적인 데이터 불균형 조건에 상응하는 성능을 가진다.

제안하는 방법이 단순한 oversampling과 데이터 증강만으로 높은 성능을 달성하지 않음을 확인하기 위해 ablation study를 수행하였다. 표 1에 baseline, oversampling, contrastive,

제안하는 방법에 대한 실험 결과를 나타내었다. 실험 결과, oversampling 기법은 coarse-grained 불균형 데이터에서는 큰 성능 향상을 보였으나, fine-grained 불균형 데이터에서는 성능 개선 효과가 크지 않았다. 반면, contrastive loss를 추가한 제안하는 방법은 coarse-grained에서 우수한 성능을 보일 뿐 만 아니라, fine-grained 조건에서 1부터 0.005의 비율까지 높은 성능을 유지하였다. 이 결과는 제안하는 방법이 단순한 데이터 분포 조정만으로는 얻을 수 없는 효과적인 feature 분리와 클래스 간 차별적 특징 학습을 달성했음을 보여준다.

4. 결론

본 연구에서는 imbalanced fine-grained image classification 문제를 해결하기 위해, oversampling과 contrastive loss를 결합한 학습 방법을 제안하였다. 제안 방법은 불균형 클래스의 oversampling과 함께 클래스 내 feature 유사성과 클래스 간 feature 비유사성을 동시에 학습하여, 데이터 불균형과 fine-grained 클래스 구분의 어려움을 함께 완화하였다. 이를 검증하기 위해 야생동물 영상 데이터셋을 대상으로 실험을 수행하였으며, 그 결과 coarse-grained와 fine-grained 조건 모두에서 기존의 oversampling 또는 cross-entropy loss 기반 방법보다 높은 분류 정확도를 달성하였다. 특히, 데이터 불균형이 심화된 조건에서도 제안 방법은 안정적인 성능을 유지하였고, 단순한 데이터 분포 조정만으로는 얻을 수 없는 효과적인 feature 분리를 실험적으로 입증하였다. 이러한 결과는 제안 방법이 데이터 불균형과 fine-grained 분류가 동시에 요구되는 다양한 영상 분류 문제에 활용될 수 있음을 시사하며, 향후 생태계 모니터링, 보전 정책 수립에 기여할 것으로 기대된다.

참고 문헌

- [1] 경의범, et al. "지리산국립공원 및 인근지역에서의 야생동물 구조 원인 분석." 국립공원연구지 12.2 (2021): 103-115.
- [2] Hassanin, Alexandre, et al. "Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes." *Comptes rendus biologies* 335.1 (2012): 32-50.
- [3] Olken, Frank. "Random sampling from databases". Diss. University of California, Berkeley, 1993.
- [4] Cui, Yin, et al. "Class-balanced loss based on effective number of samples." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.
- [5] Zhu, Jianggang, et al. "Balanced contrastive learning for long-tailed visual recognition." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.
- [6] 야생동물 활동 영상 데이터, <https://www.aihub.or.kr>, 2022.
- [7] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [8] Cubuk, Ekin D., et al. "Randaugment: Practical automated data augmentation with a reduced search space." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020.