# 강한 조명하에서 정확한 돼지 탐지를 위한 모델 앙상블

손승욱, 안한세, 이나연, 안윤호, 정용화, 박대희
고려대학교 컴퓨터융합소프트웨어학과
email: sso7199@korea.ac.kr

# Model Ensemble for Accurate Pig Detection under Strong Illumination Condition

Seungwook Son, Hanse Ahn, Nayeon Lee, Yunho An, Yongwha Chung, Daihee Park
Dept. of Computer Convergence Software, Korea University

요 약

CNN 기반 객체 탐지기의 발전으로 돈사에서 돼지 모니터링이 가능하지만, 실제 농가에서 적용하기 위해서는 영상에서 돈사의 조명에 직접 노출된 돼지들이 노출 과다 현상에 의해 탐지되지 않는 문제가 여전히 남아있다. 이러한 문제점은 싱글 모델로서는 정확도 개선의 한계가 있어, 복수개의 모델을 이용한 모델 앙상블 기법을 제안한다. 특히 본 연구에서 제안하는 영상 처리 기법을 사용하여 생성된 상호 보안적인 데이터를 통해 학습된 두 개의 TinyYOLOv4 모델을 결합하면, 돼지 객체 탐지의 정확도가 하나의 TinyYOLOv4 모델에 비하여 획기적으로 개선되었음을 확인하였다.

## 1. 서론

현재 객체 탐지 문제의 정확도를 높이기 위한 CNN 기반의 연구가 진행되고 있지만, 발표되는 많은 연구는 싱글 모델에 대한 학습 및 테스트에 초점을 맞추고 있다[1]. 하지만, 현재까지 발표된 CNN 기반 객체 탐지기중 처리 속도 대비 우수한 정확도를 제공하는 것으로 알려진 YOLOv4[2]의 COCO 데이터 셋 정확도 수준은 43%에 불과하다. 따라서 이러한 정확도 수준을 획기적으로 개선하기 위해서는 복수개의 모델을 이용한 모델 앙상블 기법의 적용이 검토되어야 한다. 사실 영상 처리 응용의 정확도 개선을 위해 복수개의 모델을 이용한 모델 앙상블 기법은 딥러닝 이전 세대부터 적용되었으며, 최근의 딥러닝 기술에 대한 앙상블 기법도 간단한 영상 분류 문제에 적용한 결과가 발표되고 있다. 또한 영상 분류 문제를 객체 탐지 문제로 확장 적용한 딥러닝 기반 앙상블 연구 결과도 최근 발표되고 있지만, 복수개의 CNN 기반 객체 탐지 모델의 다양한 조합에 의한 정확도 분석 결과는 발표되지 않고 있다.

본 연구에서는 통상의 PC 환경에서 주어진 영상이 실시간으로 처리될 수 있도록 TinyYOLOv4[2]를 두 번 적용한 모델 앙상블을 가정하며, 이러한 모델 앙상블에 사용될 두 가지 학습 방법을 제안한다. 즉, 원본 학습 데이터에 영상 전처리 기법이 적용된 데이터를 추가하여 학습하는 방법과 원본과 다른 학습 데이터를 추가하여 학습하는 방법을 고려한다. 또한, 이렇게 앙상블에 사용될 상호 보안적인 학습 데이터들의 다양한 조합에 대한 객체 탐지 정확도 실험도 수행한다.

## 2. 모델 앙상블

일반적으로 학습 모델이 서로 다르면 각 모델에서 출력되는 결과도 상이할 수 있다. 따라서, 본 연구에서는 TinyYOLOv4 에 서로 다른 학습 데이터로 학습을 시켜 생성된 객체들의 바운딩 박스들을 NMS(Non Maximum Suppression) 레벨에서 결합하는 방법을 제안한다. 구체적으로 본 연구에서는 돈사의 천장에 설치한 카메라에서 획득된 2 시간 분량의 비디오에서 움직임이 있는 키프레임을 추출하고, 이렇게 추출된 14,000장의 키프레임에 대한 정답(GT 바운딩박스)이 없는 상태를 가정한다. 또한, 돈사 모니터링 환경[3-8]에서는 하나의 돈방에 있는 돼지들이 폐사 등의 특별한 경우를 제외하면 모니터링 기간(예를 들어, 1 달) 동안 변동이 없다고 가정할 수 있다. 즉, 입력되는 키프레임에서 탐지해야 할 객체수(no_objects)를 알고 있다고 가정할 수 있다.

먼저 각 키프레임마다 두 모델 중 하나의 모델이 탐지한 바운딩 박스의 수가 프레임내 객체의 수와 동일하다면 correct 프레임이라고 간주한다. 만약 동일하지 않으면, 하나의 모델이 생성한 박스에 대하여 다른 모델의 박스 중 가장 큰 IOU(Intersection Over Union) 값을 갖는 박스를 매칭 박스로 정한다. 그리고 이 IOU 값이 두 박스가 동일하다고 간주할 수 있는 iou_thresh 값(본 연구에서는 0.7 로 설정) 보다 높으면, 두 모델의 매칭 박스가 하나의 객체를 정확히 탐지(즉, correct 박스)했다고 간주한다. 그

후 각 모델에서 correct 박스라고 간주되지 않고 남은 박스 중 높은 confidence 값을 가진 박스를 correct 박스라고 간주(전체 correct 박스 수가 no_objects될 때까지)한다. 제안 알고리즘을 요약하면 다음과 같다.

알고리즘 1, 모델 앙상블 알고리즘

```
Algorithm1. Model Ensemble Algorithm
Input: Model A, Model B, no_objects
Output: Detected Boxes C_box
Initialize:
  Detected boxes in Model A A_box
  Detected boxes in Model B B_box

if (size of A_box = no_objects) then
  return A_box as C_box
else if (size of B_box = no_objects) then
  return B_box as C_box
else {
  matched_boxes = 0
  sort A_box and B_box in descending order of
  confidence value
  for i = 1 to size of A_box do
    for j = 1 to size of B_box do
      max_iou = largest IOU of A_box[i] and
                B_box[j]
      if max_iou > iou_thresh do
        matched_boxes++
      C_box[matched_boxes] = A_box[i]
  if matched_boxes < no_objects do
    for k = matched_boxes + 1 to no_objects do
      add remaining A_box or B_box into C_box
  return C_box
}
```

## 3. 학습 데이터 생성

본 연구에서는 모델 앙상블의 효과를 극대화하기 위해 두개의 모델을 상호 보완적인 데이터로 학습하며, 크게 두 가지 학습 모델의 조합을 고려한다.

첫 번째 모델 조합은 GT 바운딩 박스가 있는 베이스라인 학습 데이터에 CLAHE(Contrast-Limited Adaptive Histogram Equalization) 영상 처리 기법을 적용한 학습 데이터를 사용하여 학습한 모델 조합을 제안한다. 특히 CLAHE의 파라미터 값을 달리 설정함으로써, 전배경 분리와 돼지의 질감을 각각 강조하는 영상 처리 결과를 얻을 수 있다.

두번째 모델 조합은 기존 학습 데이터에, 추가적인 학습 데이터를 이용하여 데이터 증강을 한 후 학습시킨 모델 조합을 제안한다. 일반적으로 데이터 증강은 기존 학습 데이터에 영상 처리 기법 등을 적용하여 학습 데이터의 수를 증가시키는 것이다. 하지만 본 연구에서는 두번째 학습 모델에 대하여 이러한 데이터 증강 외에 다른 학습 데이터를 추가하는 것으로 데이터 증강을 수행하였다. 이 학습 모델은 기존 학습 데이터와 추가된 학습 데이터의 겹치는 특징을 학습하여 그 부분을 보다 정확하게 탐지하기 위해 제안한다.

## 4. 실험 결과

본 실험은 충청북도 청주시에 위치한 충북대학교 내에 9마리 돼지(110 ~ 115일령, 90 ~ 98kg, Duroc × Landrace × Yorkshire)가 활동하는 돈사를 모니터링하면서 수행되었다. 영상 데이터 촬영을 위해 돈사의 중앙을 기준으로 3.2m 높이의 천장에 Intel® RealSense D435 카메라를 설치하였고, 이를 통해 1280 × 720 해상도의 영상 데이터를 획득하였다. 그림 1은 획득한 영상 데이터의 예시를 보여준다. 또한, one-stage 딥러닝 구조로 빠른 객체 탐지에서 널리 이용되는 YOLOv4 및 TinyYOLOv4[2]를 사용하였으며, Ubuntu 18.04 LTS OS, Intel(R) Core(TM) i7-9700 CPU, NVIDIA RTX 2080TI GPU, 16GB RAM 환경에서 실험을 진행하였다.

학습 데이터는 873장의 돼지 학습 데이터를 사용하였으며, 영상 전처리 모델의 경우 CLAHE를 적용함으로 인해 각 모델당 873장이 추가되어 1,746장에서 학습이 진행되었다. 데이터 증강에서는 1,000장이 추가되어 1,873장으로 학습을 진행하였다. 테스트 데이터는 돈사에서 촬영한 9마리의 돼지가 움직이는 동영상에서 추출한 14,000장의 키프레임을 이용하여 진행하였으며, 탐지된 박스의 수가 9개가 아닌 프레임을 에러 프레임으로 간주하였다.

표 1은 영상 전처리가 적용된 학습 데이터로 테스트한 결과를 나타내며, 영상 전처리 모델 A는 명암 대비를 강조한 CLAHE 설정이 적용이 된 학습 데이터를, 영상 전처리 모델 B는 객체내의 질감을 강조한 CLAHE 설정이 적용이 된 학습 데이터를 의미하고 이는 그림 2-3에서 표현되고 있다. 실험 결과, 표 1에서 영상 전처리의 학습 모델을 모델 앙상블을 통해 결합할 때 에러 프레임이 TinyYOLOv4의 4,886장에서 3,652장으로 대폭 감소함을 확인할 수 있었다. 또한, 이러한 에러수는 YOLOv4의 싱글 모델(3,976장) 보다 적음을 확인하였다.

그림1. 돈사 모니터링 원본 영상

그림2. 그림1에 영상 전처리 A를 적용한 영상

그림3. 그림1에 영상 전처리 B를 적용한 영상

본 연구에서는 데이터 증강 학습 모델에 대하여 이러한 데이터 증강 외에 다른 학습 데이터를 추가하는 것으로 데이터 증강을 수행하였다. 이 학습 모델은 기존 학습 데이터와 추가된 학습 데이터의 겹치는 특징을 학습하여 그 부분을 보다 정확하게 탐지하기 위해 제안한다. 이러한 데이터 증강을 위하여 본 연구에서 사용된 학습 데이터는 세포 데이터[9]이며 그림 4에서 나타내고 있다. 세포 데이터는 배경과 객체의 분리가 명확하고 객체들 사이에서 겹치는 부분이 많이 존재하는 학습 데이터이다. [10]와 같이 객체 탐지기가 세포의 윤곽선을 학습하게 하여 돼지가 서로 겹치는 부분에서의 오류를 최소화하고자 하였다. 표 2는 추가 학습 데이터를 이용한 테스트 결과를 나타내며, 데이터 증강 C 모델은 기존 학습 데이터와 추가 데이터를 같은 클래스로, 데이터 증강 D 모델은 다른 클래스라고 분류하고 학습한 모델을 의미한다. 실험 결과, 싱글 모델에서는 베이스라인 모델에 비하여 에러가 크게 증가하였지만, 모델 앙상블을 통해 베이스라인 모델에 비하여 에러가 감소함을 확인할 수 있었다. 그러나 표 1에서 보여준 영상 전처리에 의한 앙상블 효과에 비하여 효과적이지 않음을 확인하였다.

그림 4. 세포 dataset 예시

표 3은 표1과 표2에 나타난 학습 모델의 조합을 모델 앙상블 한 결과를 나타내며 각각 표1의 두 개의 학습 모델과 표2의 두 개의 학습 모델을 조합한 모델을 의미한다. 실험 결과 각 표1과 표2의 싱글 모델 에러 프레임 수보다 표3의 모델 앙상블 에러프레임 수가 감소 하였음을 확인하였다. 또한, 영상 전처리 B를 포함한 모델 앙상블 결과가 영상 전처리 A를 포함한 모델 앙상블 결과보다 에러 프레임 수가 더 적음을 확인 하였다. 표2의 모델 앙상블 효과에 비해서 표3의 모델 앙상블 효과가 더 효과적이며, 표3의 모델 앙상블 효과에 비해서 표1의 모델 앙상블 효과가 더 효과적임을 확인 하였다.

표 1. 상이한 영상 전처리로 TinyYOLOv4에서 학습된 두개 모델의 앙상블 효과(테스트 프레임 수: 14,000, Confidence 0.9 기준)

| 모델 | | 에러 프레임 수 |
|---|---|---|
| 싱글 모델 | 베이스라인 | 4,886 |
| | 영상 전처리 A | 5,833 |
| | 영상 전처리 B | 4,855 |
| 모델 앙상블 | | 3,652 |

표 2. 상이한 데이터 증강으로 TinyYOLOv4에서 학습된 두개 모델의 앙상블 효과(테스트 프레임 수: 14,000, Confidence 0.9 기준)

| 모델 | | 에러 프레임 수 |
|---|---|---|
| 싱글 모델 | 베이스라인 | 4,886 |
| | 데이터 증강 C | 7,070 |
| | 데이터 증강 D | 6,582 |
| 모델 앙상블 | | 4,307 |

표 3. 상이한 데이터 증강으로 TinyYOLOv4에서 학습된 두개 모델의 앙상블 조합의 효과(테스트 프레임 수: 14,000, Confidence 0.9 기준)

| 모델 | | 에러 프레임 수 |
|---|---|---|
| 모델 앙상블 | 영상 전처리 A + 데이터 증강 C | 4,908 |
| | 영상 전처리 A + 데이터 증강 D | 5,016 |
| | 영상 전처리 B + 데이터 증강 C | 4,089 |
| | 영상 전처리 B + 데이터 증강 D | 4,162 |

## 5. 결론 및 향후 연구

본 연구에서는 두개의 모델을 결합한 모델 앙상블을 통해, 하나의 모델에서 탐지 못하는 객체를 다른 모델의 결과를 결합하여 보완할 수 있는 방법을 제안하였다. 특히, 상이한 영상 전처리 및 데이터 증강으로 학습된 두개 모델을 NMS를 통한 후처리 레벨에서 결합함으로써, 두개의 모델이 상호 보완적인 결과를 낼 수 있도록 고려하였다. 돈방에서 획득한 비디오 데이터로 실험한 결과, 싱글 모델의 오류 프레임 수를 크게 개선할 수 있음을 확인하

였다. 또한, 모델 앙상블의 효과로 인해 전체적으로 영상 전처리 B가 영상 전처리 A보다 정확도에 더 효과적임을 확인하였다. 향후 연구로서 다른 싱글 모델을 연구하여 모델 앙상블에 더 좋은 효과를 부여 할 수 있는 연구를 진행할 계획이다.

## 감사의 글

본 연구는 2018년도 교육부의 재원으로 한국연구재단의 기초연구사업(2018R1D1A1A09081924)과 2019년도 과학기술정보통신부의 재원으로 한국연구재단-현장맞춤형 이공계 인재양성 지원사업(2019H1D8A1109907)의 지원으로 수행된 연구결과임.

## 참고 문헌

[1] L. Liu, et al., “Deep Learning for Generic Object Detection: A Survey,” *International Journal of Computer Vision*, Vol. 128, pp. 261-318, 2020.

[2] A. Bochkovskiy, C. Wang, and H. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *arXiv preprint arXiv:2004.10934*, 2020.

[3] S. Matthews, et al., “Early Detection of Health and Welfare Compromises through Automated Detection of Behavioural Changes in Pigs,” *The Veterinary Journal*, Vol. 217, pp. 43-51, 2016.

[4] J. Lee, J, L. Jin, D. Park, and Y. Chung, “Automatic Recognition of Aggressive Behavior in Pigs using a Kinect Depth Sensor.” *Sensors*, Vol. 16, pp. 631, 2016.

[5] J. Kim, Y. Chung, Y. Choi, J. Sa, H. Kim, Y. Chung, D. Park, and H. Kim, “Depth-based Detection of Standing-Pigs in Moving Noise Environments,” *Sensors*, Vol. 17, pp. 2757, 2017.

[6] J. Sa, Y. Choi, H. Lee, Y. Chung, D. Park, and J. Cho, “Fast Pig Detection with a Top-View Camera under Various Illumination Conditions,” *Symmetry*, Vol. 11, pp. 266, 2019.

[7] Y. Choi, J. Sa, Y. Chung, and D. Park, “보온등 환경에서 깊이 정보를 이용한 돼지 탐지,” 한국정보처리학회 춘계학술대회논문집, pp. 693 - 695, 2016.

[8] J. Seo, J. Sa, Y. Choi, Y. Chung, D. Park, and H. Kim, “A YOLO-based Separation of Touching-Pigs for Smart Pig Farm Applications,” In Proceedings of the 21st International Conference on Advanced Communication Technology (ICACT), pp. 395-401, 2019.

[9] K. Sofiiuk, et al., “AdaptIS: Adaptive Instance Selection Network,” In Proc. of ICCV, pp. 7355-7363, 2019.

[10] P. Dollar, Z. Tu, and S. Belongie, “Supervised Learning of Edges and Object Boundaries,” In Proc. of CVPR, pp. 1964-1971, 2006.