

VGG-16 모형 및 CAM 기법을 이용한 개의 피부병 이미지 분류 및 관심영역 도출

이병우¹⁾ · 류재석²⁾ · 정세민³⁾ · 문보미⁴⁾ · Yang Sok Kim⁵⁾

¹²³⁴⁵계명대학교 경영정보학과

요약

반려동물을 키우는 가구가 증가함에 따라 반려동물 의료에 대한 관심도 커지고 있다. 또한 머신러닝 과 딥러닝의 발전으로 이미지 분야에서 식별 및 분류에 대한 연구가 많이 진행되고 있다. 본 논문에서는 CNN(convolution neural network) 기반의 VGG-16모형을 수정한 수정된 VGG-16모형을 이용하여 반려동물 중 개의 피부병을 분류하였다. 결과 비교는 Global Average Pooling과 Global Max Pooling를 비교했으며 각각 정확도(Accuracy)는 0.919, 0.926으로 Global Max Pooling의 성능이 약간 우수했다. CAM(class activation map)기법을 이용하여 수정된 VGG-16 모형이 이미지의 어느 부분에 관심영역을 두고 분류하였는지 히트맵으로 나타내었으며, 관심영역을 살펴본 결과, 정확하게 피부병이 있는 부위를 보고 분류가 되지는 않았다. 그 이유로는 개의 종별로 털의 색상과 길이, 피부의 주름, 피부의 점의 유무 등의 차이 때문인 것으로 보였으며, 피부병의 경우 특성상 털이 빠져있기 때문에 피부병이 있는 부위에 모형이 특징을 두는 것이 아닌 털이 없는 부분에 관심영역이 잡혀 있음을 확인하였다. 따라서 개의 피부병 분류를 위해서는 개의 종에 따라 데이터를 나누어 따로 모델에 학습해야 할 것으로 보이며, 피부병 부위를 마스크(masking)처리를 통해 피부병이 있는 부위를 먼저 식별한 후 분류를 진행한다면 더 효과적인 분류를 할 수 있을 것이다.

주요용어: 개의 피부병, 관심영역, 분류, CAM, CNN, VGG-16,

1. 서론

현재 검역본부의 통계를 보면, 반려동물을 키우는 가구가 한국에서 2019년 1,418만명에 육박하며 이에 대한 비율도 계속해서 높아가고 있는 추세이다. 반려동물 의료

1) (42601) 대구광역시 달서구 달구벌대로 1095, 계명대학교 경영정보학과 석사과정.

2) (42601) 대구광역시 달서구 달구벌대로 1095, 계명대학교 경영정보학과 학사

3) (42601) 대구광역시 달서구 달구벌대로 1095, 계명대학교 경영정보학과 학사

4) (42601) 대구광역시 달서구 달구벌대로 1095, 계명대학교 경영정보학과 학사

5) 교신저자: (42601) 대구광역시 달서구 달구벌대로 1095, 계명대학교 경영정보학과 교수.

E-mail: yangsok.kim@kmu.ac.kr

에 대한 관심도 지속적으로 늘고 있으며, 반려동물에 대한 인식 또한 평범한 동물이 아닌 가정에서 함께 생활하는 수준이 되고 있다. 한편, 머신러닝(machine learning)과 딥러닝(deep learning)의 발전으로 눈으로 보고 진단할 수 있는 CT영상 또는 X선 사진을 학습을 통해 분류하는 연구(Liu 등 2020) 및 CT영상의 잡음을 제거하는 연구(Kim 등 2020)가 있었으며 직관적으로 증상을 눈으로 볼 수 있는 사람의 피부병의 발진 또는 피부암에 대한 식별 및 분류연구(Li 와 Shen 2018)도 진행된 바 있다.

본 논문에서는 사람의 피부병의 발진 또는 피부암을 식별하는 모델을 토대로 반려동물 중 개의 피부병을 분류할 수 있도록 데이터를 수집하여 실험을 진행하였다. 분류는 이미지의 패턴 및 특징을 학습할 수 있는 CNN(convolution neural network)기반의 VGG-16모델을 이용하였으며, 이 실험을 통해 개의 피부병을 식별을 진행해 보고, 문제점 및 한계점을 도출하고자 한다. 이를 위해 VGG-16모델을 수정하여 CAM(class activation map)기법을 통해 분류모델의 결과가 이미지의 어느 부분에 관심영역을 가지고 분류했는지 도출하고자 한다.

2. 연구 방법

2.1 수정된 VGG-16 모형

분류에 이용된 모델은 VGG-16 모형의 구조를 이용하였으며, CNN(Convolution Neural Network)기반의 모형이며 다른 모형에 비해 합성곱 Layer에 3x3필터를 적용하여 파라미터(parameter) 수를 많이 줄인 모형이다(Karen 와 Andrew 2015). 합성곱 층은 총 5개의 층을 이용하며 각층은 Relu 활성화 함수를 이용한다. ReLu의 식은 다음 식 2.1과 같으며 0보다 작은 값들은 0으로 처리하게 된다.

(2.1)

$$ReLU(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$$

1개의 층을 통과 후 MaxPooling을 2x2 커널(kernel)과 stride를 2로 설정하여 이용하며 이미지의 특징(Feature)에서 가장 큰 신호에 대한 정보만 남긴 후 그 크기를 반감 시키고 채널은 늘리도록 구성이 되어 있으며 이 과정을 통해서 다음 합성곱 층(convolution layer)으로 정보를 전달한다.

본 논문에서 이용한 모형은 기존에 제안된 VGG-16 모형에서 마지막 판별층(fully connected layer)을 제외시키고 마지막 합성곱 층(convolution layer)에서 채널을 1024개로 늘린 후 Global Average Pooling 또는 Global Max Pooling을 통해 1차원 벡터에 이미지의 Feature를 반영하도록 하였다. 이 Global Average Pooling은 각 이미지의 Feature map상의 노드값의 평균을 도출한다. Global Max Pooling은 각 이미지의 Feature map상의 노드값의 최대값을 도출한다. 따라서 크기가 감소된 Feature map을 도출할 수 있으며 파라미터의 수가 줄어들어 모델의 학습 속도가 빨라지며, 학습의 과적합도 줄일 수 있다.

수정된 VGG-16 모형은 다음의 Cross entropy 손실함수를 줄이는 방향으로 학습이 진행되며 식은 다음과 같다.

(2.2)

$$-\sum_{i=1}^n t_i' P(s_i)$$

여기서 t 와 s 는 각각 실제값과 예측값을 뜻하고 i 는 각 label을 의미한다. 이를 최소화하기 위해서 Adam 알고리즘을 통해 학습을 진행한다.

마지막으로 Global Pooling층의 Feature를 이용하여 softmax층을 통해 정상 또는 피부병으로 예측하도록 구성 되어있으며, Global Average Pooling과 Global Max Pooling을 각각 적용하여 결과를 비교해 볼 것이다. 수정된 VGG-16 모형의 구성은 다음 Figure 2.1과 같다.

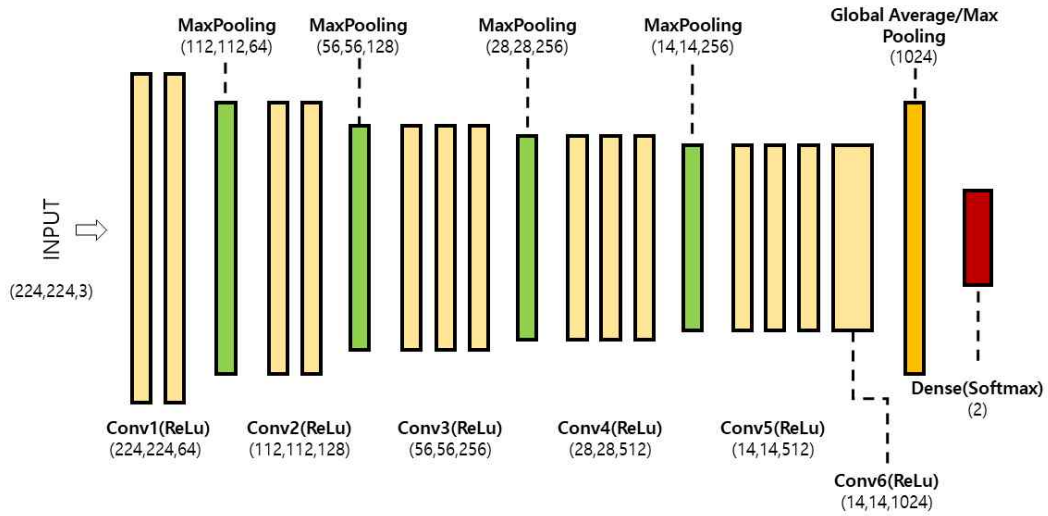


Figure 2.1 VGG-16 Model

2.2 CAM기법을 이용한 모형의 관심영역 도출

CAM(Class Activation Map)기법은 분류 모형의 결과가 이미지의 어느 부분이 분류 결정에 큰 영향을 주었는지 분석하고자 할 때 쓰는 기법이며, 마지막 판별 레이어에서 가지는 가중치값을 합성곱 층과 Pooling층을 거친 $n*n$ 행렬에 곱하면 판별식에서 어느 부분이 큰 값을 가졌는지 알 수 있다. 과정을 식으로 나타내면 다음과 같다.

(2.3)

$$f_k(x, y)$$

이는Figure 2.2의 Conv6의 k 번째 채널의 값들 중 (x, y) 에 위치한 값을 뜻한다. Global Average Pooling을 거친 값을 식으로 나타내면 다음 (2.4)식과 같다.

$$F^k = \sum_{x,y} f_k(x,y) \quad (2.4)$$

분류되는 값 c 에 대해 softmax층으로 입력되는 값은 다음 식(2.5)과 같다. w_k^c 가 k 번째 채널과 분류되는 값 c 에 해당하는 값이라면 w_k^c 는 예측되는 값인 c 에서 F_k 만큼 중요하다는 것을 나타내며 S_c 는 F_k 에 각 채널별 값을 포함한다.

$$\begin{aligned} S_c &= \sum_k w_k^c \sum_{x,y} f_k(x,y) \\ &= \sum_{x,y} \sum_k w_k^c f_k(x,y) \end{aligned} \quad (2.5)$$

M_c 는 c 에 대한 CAM(class activation map)값이며 다음 (2.7)식을 통해 도출된다.

$$M_c(x,y) = \sum_k w_k^c f_k(x,y) \quad (2.7)$$

따라서 S_c 를 다시 나타내게 되면 (2.8)식과 같다.

$$S_c = \sum_{x,y} M_c(x,y) \quad (2.8)$$

해당 채널의 Feature값에 가중치를 곱한 수치를 이용하여 중요도로 해석하는 방식이며, 히트맵을 $n*n$ 행렬의 크기로 만들어 시각적으로 결정에 큰 영향을 준 이미지의 위치를 볼 수 있다. 그려진 히트맵은 이미지에서의 위치정보를 그대로 반영하고 있기 때문에 가장 높은 값을 가지는 위치가 분류에 있어서 가장 영향력이 높은 부분이라고 해석할 수 있다(Zhou 등 2016).

2.3 모형 구현

수정된 VGG-16 모형 및 CAM(class activation map)은 Python 언어를 이용하였으며, 딥러닝 구현을 위한 프레임워크인 텐서플로우(Tensorflow)와 케라스(Keras)를 이용하여 모형을 구현하였으며 구현 과정은 다음 Figure2.2와 같다.

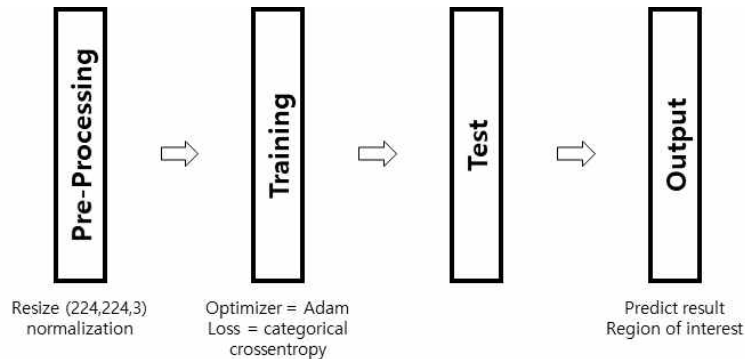


Figure 2.2 process configuration

실험 모형의 과정은 다음 Figure 2.2와 같으며 Figure 2.1의 수정된 VGG-16 모형에 이미지와 각 이미지에 맞는 Label이 Input되며 손실함수는 categorical cross entropy에 의해 계산되고 최적화 알고리즘으로는 Adam 알고리즘을 이용하였다. 학습의 epoch는 20, batch size는 10으로 설정하여 모형 학습을 진행하였다. 마지막 Output으로는 정상 및 비정상에 대한 예측 결과값과 모형의 분류결과 관심영역 결과가 출력된다.

3. 실험 설계

3.1 데이터 수집 및 준비

분류모형에 이용할 데이터는 Python을 이용하여 Google 검색결과 중 해당하는 이미지를 크롤링(crawling)하여 준비하였으며, 개의 피부병이 아닌 정상인 데이터는 피부병을 앓지 않은 개의 이미지에서 각각 다리와 피부에 해당하는 부분을 잘라서 이용하였다. 개의 피부병은 다른 피부병보다 눈에 잘 띄는 피부병들을 대상으로 하였다.

피부병에 해당하는 사진은 Acral lick dermatitis 사례 60장, allergic 사례 157장, impetigo 사례 82장, Ringworm 사례 94장으로 피부병에 해당하는 사진은 총 393장을 이용했다. 정상에 해당하는 사진은 355장으로 개의 종 구분 없이 털과 피부, 발에 해당하는 데이터 셋을 구성하였으며 데이터 셋의 예시는 다음 Figure 3.1과 같다. 총 4가지의 피부병을 모두 Abnormal로, 나머지 정상에 해당하는 사진은 Normal로 이진분류로 다룰 수 있도록 데이터 셋을 구성하였다.

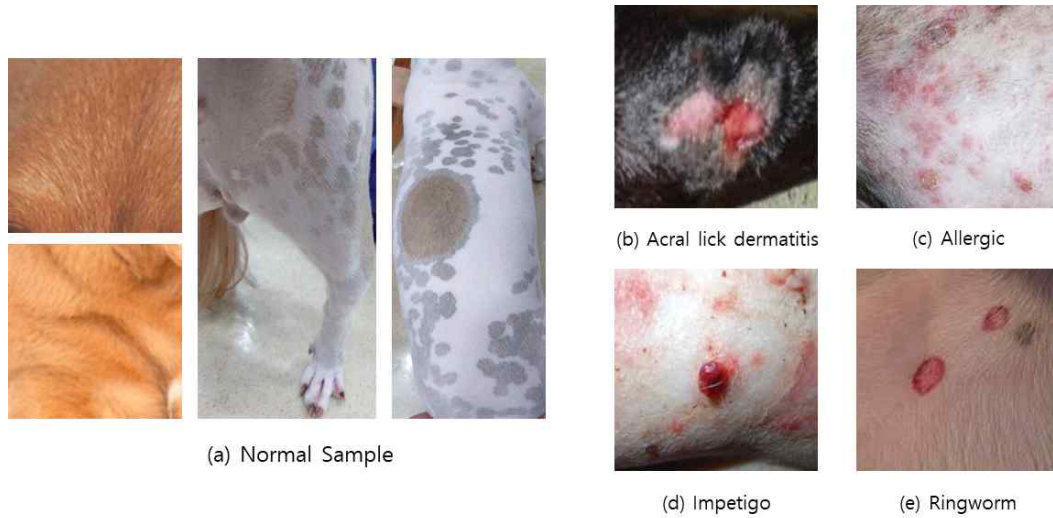


Figure 3.1 (a) Normal Sample, (b) Abnormal Sample – Acral lick dermatitis, (c) Abnormal Sample – Allergic, (d) Abnormal Sample – Impetigo, (e) Abnormal Sample – Ringworm

학습용 데이터는 정상 284장과 비정상 315장 총 599장으로 구성하였으며, 테스트용 데이터는 정상 71장과 비정상 78장 총 149장으로 구성하였다. 이미지에 대한 전처리는 크기를 모두 동일하게 224x224로 맞춰주었으며 각 이미지의 픽셀값을 0~1 사이 값으로 표준화 시켜주었다.

3.2 Global Pooling 비교

Figure 2.1의 모형에서 Global Pooling 층을 각각 Global Average Pooling과 Global Max Pooling을 이용하여 분류 정확도(accuracy) 및 CAM(class activation map)의 관심영역을 비교하였다. 모형의 채널수는 모두 동일하게 1024로 설정하였으며, 학습에 관한 파라미터 또한 동일하게 epoch 20, batch size 10으로 설정하였다. 관심영역의 비교는 모델이 테스트 데이터에서 피부병으로 맞춘 사례와 오분류한 사례를 비교할 수 있도록 하였다.

4. 실험 결과

4.1 분류 성과 지표

분류 성과는 테스트 데이터 정상 71장과 비정상 78장에 대한 정확도(accuracy)를 측정하였다. 정확도(accuracy)는 가장 직관적으로 모델의 성능을 나타낼 수 있는 평가 지표이며 식은 다음과 같다.

$$(4.1)$$

$$(Accuracy) = \frac{TP + TN}{TP + FN + FP + TN}$$

TP는 실제 정상을 정상으로 예측한 건수, FN은 실제 정상을 비정상으로 예측한 건수, FP는 실제 비정상을 정상으로 예측한 건수, TN은 실제 비정상을 비정상으로 예측한 건수로 나타낸다.

Global Average Pooling을 이용하였을 때 분류 성능은 정확도(accuracy) 0.919를 기록하였으며 Figure 4.1은 학습 및 테스트의 Epoch당 Loss와 정확도(accuracy)이다.

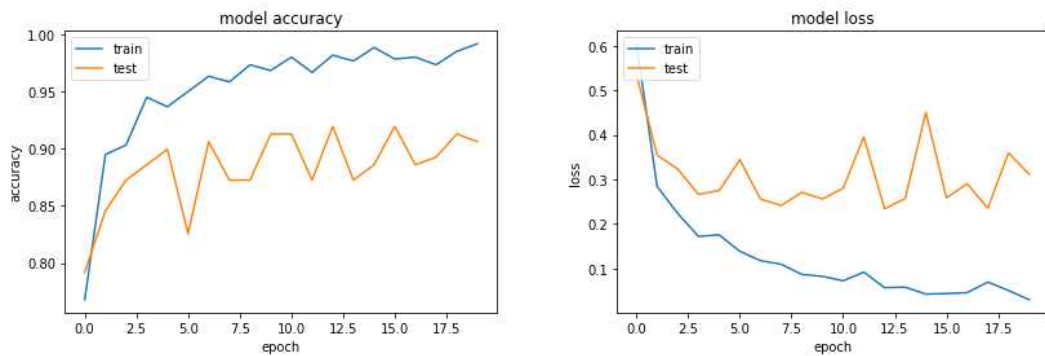


Figure 4.1 Global Average Pooling' s accuracy and loss graph per epoch

Global Max Pooling을 이용하였을 때 분류 성능은 정확도(accuracy) 0.926을 기록하였으며 Figure 4.2은 학습 및 테스트의 Epoch당 Loss와 정확도(accuracy)이다. Global Max Pooling을 이용하였을 때는 3 epoch부터 학습이 진행되어도 정확도와 loss가 개선되지 않는 것이 확인된다.

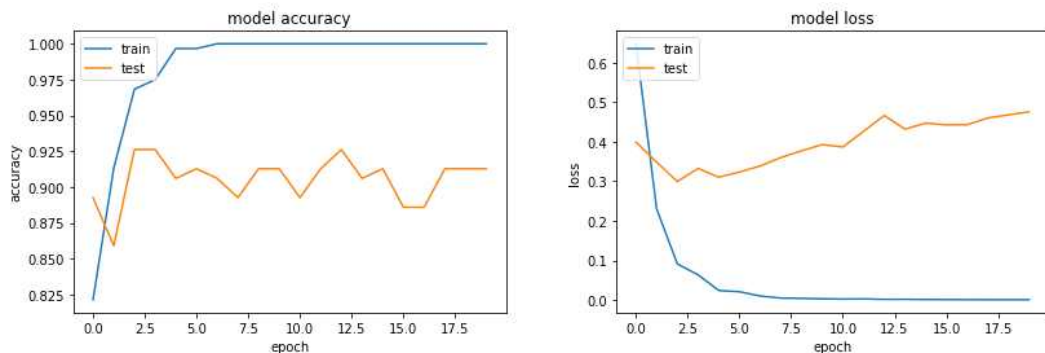


Figure 4.2 Global Max Pooling' s accuracy and loss graph per epoch

4.2 관심영역 도출

수정된 VGG-16모형에 Global Average Pooling과 Global Max Pooling을 이용한 분류 결과에 CAM(class activation map)기법을 이용하여 관심영역 히트맵을 도출하였다. Global Average Pooling과 Global Max Pooling의 관심영역 비교는 테스트 데이터 중 피부병을 피부병으로 맞춘 경우와 오분류가 발생한 정상을 피부병으로 예측한 경우, 피부병을 정상으로 예측한 경우를 비교하였다.

Figure 4.3과 4.4는 Global Average Pooling과 Global Max Pooling의 피부병을 맞춘 결과에 대한 관심영역을 살펴본 결과로, 대부분 피부병이 존재하는 영역에 관심영역이 잡혀있으나 부정확하게 피부병을 인식한 샘플도 존재한다.

Figure 4.5과 4.6는 Global Average Pooling과 Global Max Pooling의 오분류 중 정상샘플을 피부병으로 예측한 결과에 대한 관심영역을 시각화 한 것으로, 대개 피부에 점이 있거나 주름진 샘플, 털의 색상이 차이나는 부분 때문에 오분류가 나타났음을 알 수 있다.

Figure 4.7와 4.8은 Global Average Pooling과 Global Max Pooling의 오분류 중 피부병 샘플을 정상으로 예측한 결과에 대한 관심영역을 시각화 한 것으로 피부병인 부분보다 털이 존재하는 부분에 관심영역이 더 잡혀 있음을 확일 할 수 있다.

종합적으로 보았을 때, Global Average Pooling과 Global Max Pooling의 분류결과와는 약간의 차이가 있었으나 관심영역은 비슷하게 형성이 된 것을 확인할 수 있으며 오분류는 개의 피부의 점, 털의 특징 등의 원인 때문에 발생하는 것으로 보인다.

Classification of Dog' s Skin Disease Images and Extracting Region of **277**
Interest Using VGG-16 Models and CAM Techniques

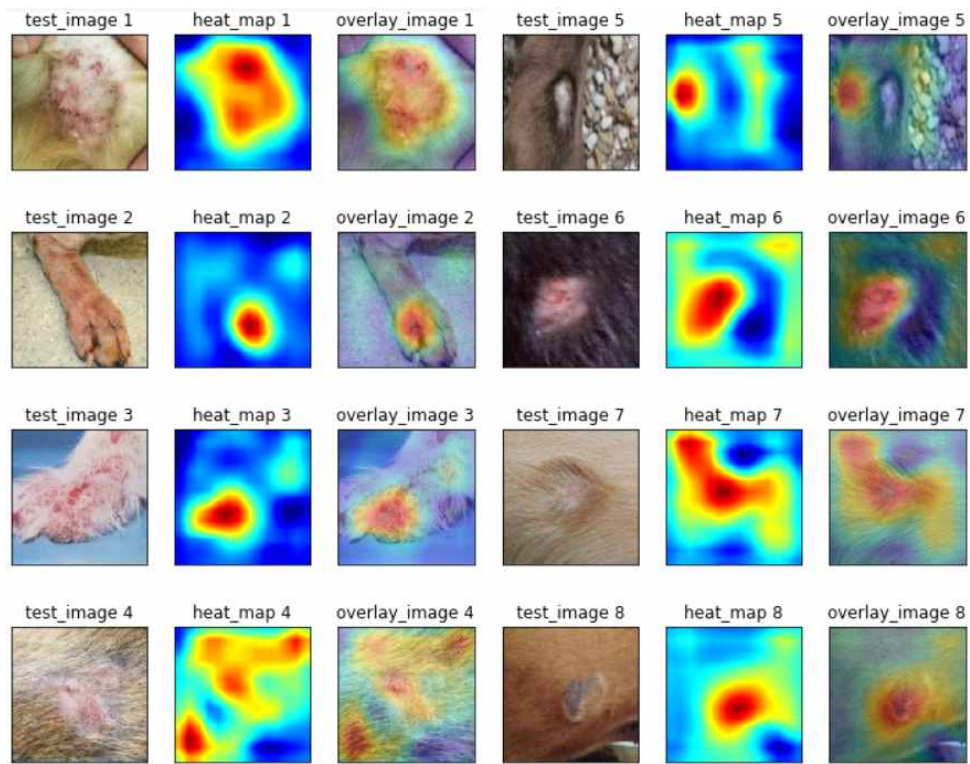


Figure 4.3 Global Average Pooling' s abnormal correct result

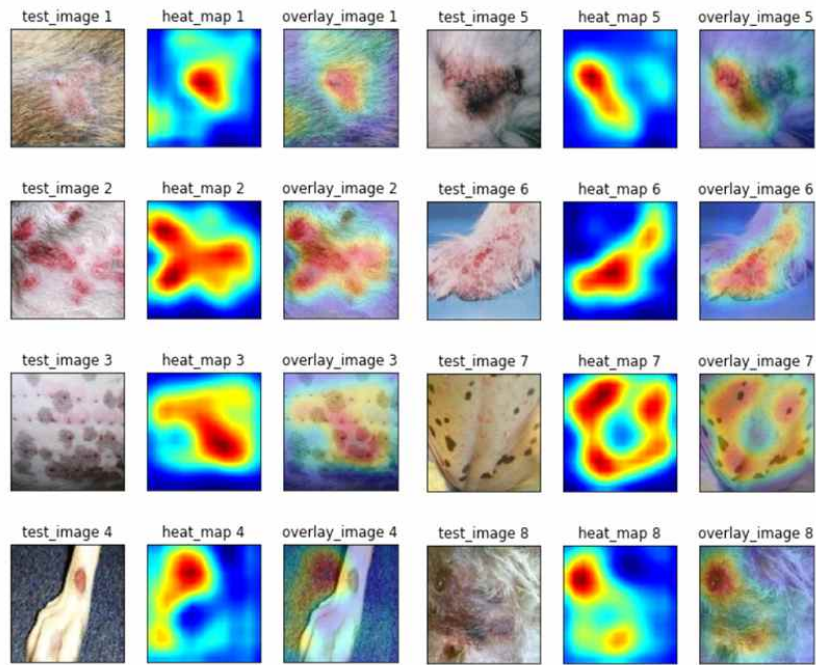


Figure 4.4 Global Max Pooling's abnormal correct result

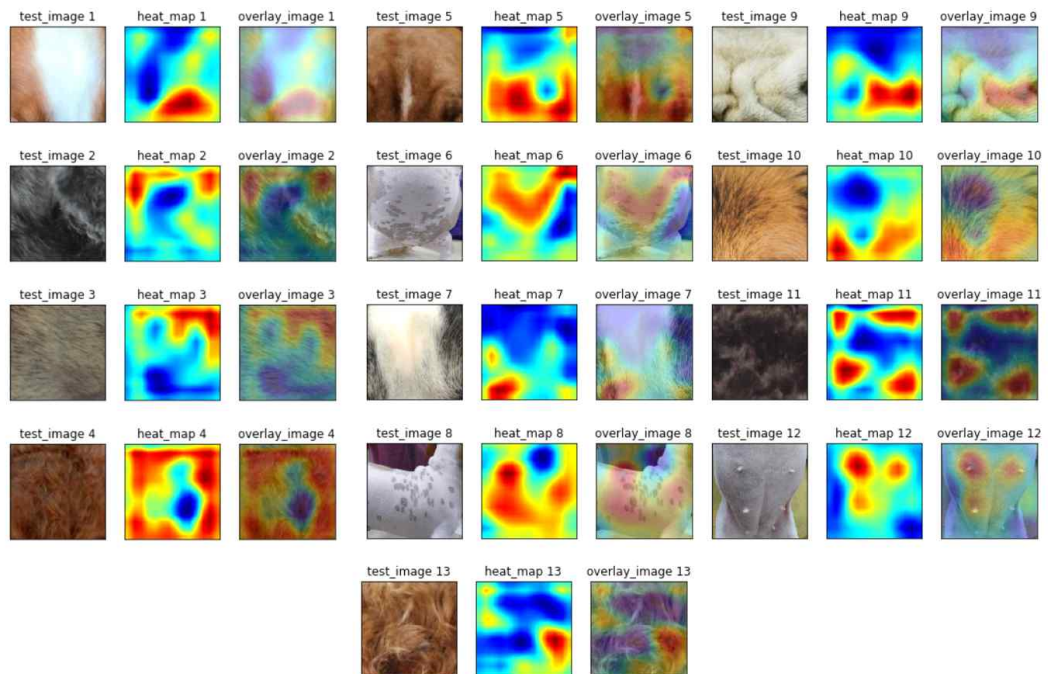


Figure 4.5 Global Average Pooling' s abnormal wrong result

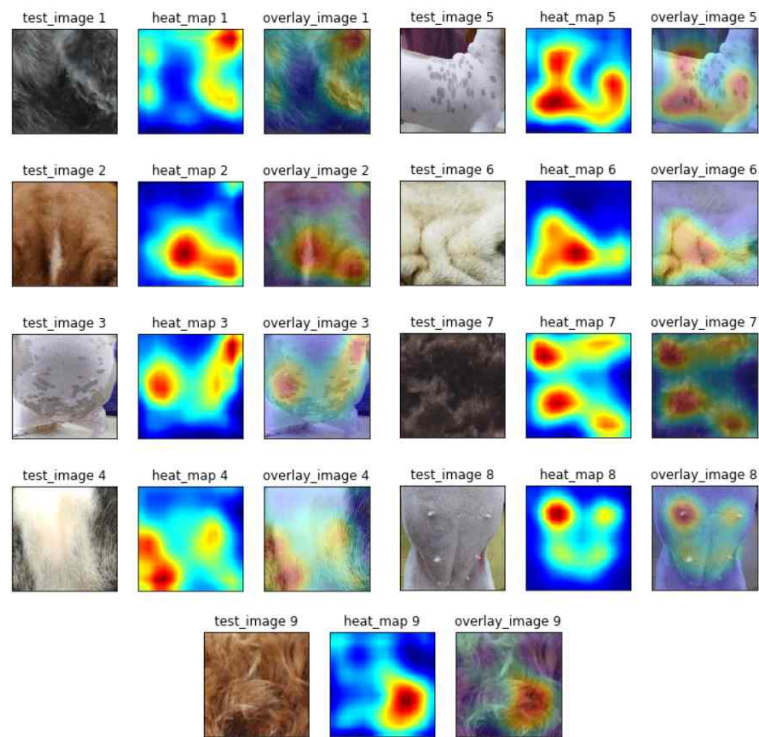


Figure 4.6 Global Max Pooling' s abnormal wrong result

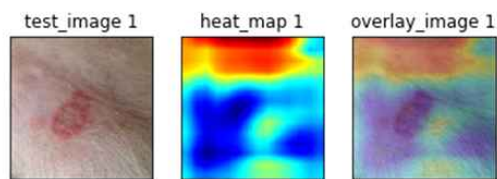


Figure 4.7 Global Average Pooling' s normal wrong result

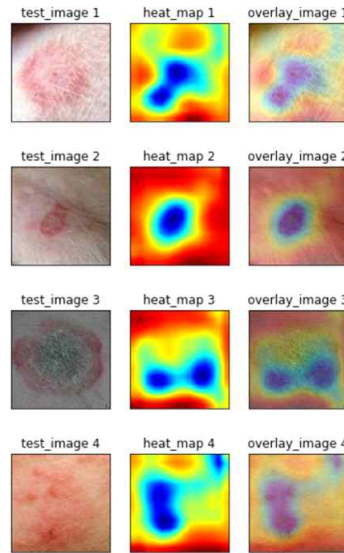


Figure 4.8 Global Max Pooling' s normal wrong result

5. 결론 및 한계점

이전의 이미지 분석을 다루는 딥러닝 방법론에서 항상 대두되는 것이 모형이 어떤 부분을 보고 분류는 했는지에 대한 것이었다. 하지만 CAM(class activation map)기법이 나오면서 분류 전 단계의 가중치(weight)를 이용하여 모형의 관심영역을 확인할 수 있게 되었다. 이를 이용하여 분류결과와 해석과 이용하고자 하는 데이터 또는 도메인에서 모형이 어떻게 인식되는지 파악하여 데이터의 확인절차로도 이용할 수 있다.

본 논문에서는 개의 피부병 데이터와 정상 데이터를 수정된 VGG-16모형을 이용하여 피부병에 대한 분류를 진행해보았다. 기존 VGG-16모형의 구조와는 다르게 판별층(fully connected layer)부분을 제거하고 이미지 특징(feature)을 1차원 벡터로 나타내기 위해 Global Average Pooling을 이용하였다. 분류 결과는 Global Average Pooling을 이용했을 때 정확도(accuracy)는 0.919, Global Max Pooling을 이용했을 때 정확도(accuracy)는 0.926이었으며 Global Average Pooling을 이용했을 때 보다 성능이 약간 좋았다.

모형의 관심영역을 도출하기 위해서 CAM(class activation map)기법을 이용하였으며 적용 과정은 Figure 2.1의 Conv6의 14x14x1024 크기의 이미지 특성(feature)을 Global Average Pooling을 이용하여 길이가 1024인 벡터화 시켜주었으며, 마지막 판별층에서 2개의 Class를 예측하게 된다. Global Average Pooling과 Class를 연결하는 가중치 매트릭스(weight matrix)는 1024x2 크기의 매트릭스이며 가장 높은 확률을 가지는 Class를 선택하여 1024x1의 가중치 벡터(weight vector)를 구한 후 이미지 원본사이즈로 맞춰주어 가중치 벡터(weight vector)를 곱하는 방식으로 모형의 관심영역을 도출하였다.

도출 결과 개의 종에 따라 털의 색 및 길이, 피부병 발생 부위가 다른 점과 피부병의 특성상 털이 빠지는 특성 때문에 털이 짧거나 없는 샘플에 대해서 오분류가 나타남을 알 수 있었다. 따라서 더 정확한 분류가 이루어지기 위해서는 개의 종별로 데이터를 따로 선별하여 모델링하거나 피부병이 있는 부위를 마스킹(masking)하여 피부병의 위치를 먼저 선별한 후 분류하는 방식이 필요해 보인다. 또한, 개 또는 반려동물들에 대한 데이터 샘플의 부족으로 양질의 데이터를 더 많이 확보해야 할 것이다.

향후 연구에서는 데이터를 추가 확보하여 각 피부병에 대한 다중 분류를 실시할 예정이며 본 논문에서 관심영역 확인 결과를 반영하여 개의 종을 분류하거나 피부병 부분에 대한 마스킹 처리를 통해 인식률을 높여 분류하고자 한다.

참고문헌

- B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba (2016). Learning Deep Features for Discriminative Localization. *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2921–2929.
- Kim. J. H., Song. J. H., and Lim. D. H (2020). CT image denoising using inception model. *Journal of the Korean Data & Information Science Society*, **31**, 487–501.
- K. Simonyan, A. Zisserman (2014). Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*,
- Kingma, D. and Ba, J. (2015). Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 1–15.
- Villan, A. F. (2019). Mastering OpenCV 4 with Python: : A practical guide covering topics from image processing, Augmented reality to deep learning with OpenCV 4 and Python 3.7, Packt Publishing.
- T. Y. Hsiao, Y. C. Chang, H. H. Chou, C. T. Chiu (2018). Filter-based deep-compression with global average pooling for convolutional networks. *Journal of Systems Architecture*, **95**. 9–18.
- Y. Li, L. Shen (2018). Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network. *Sensors*, **18**, 556.
- Y. Liu, A. Jain, C. Eng, D. H. Way, K. Lee, P. Bui, K. Kanada, G. O. Marinho, J. Gallegos, S. Gabriele, V. Gupta, N. Singh, V. Natarajan, R. Hofmann-Wellenhof, G. S. Corrado, L. H. Peng, D. R. webster, D. Ai, S. J. Huang, Y. Liu, R. C. Dunn and D. Coz (2020). A deep learning system for differential diagnosis of skin diseases. *Nature Medicine*, **26**. 900–908.

Classification of Dog' s Skin Disease Images and Extracting Region of Interest Using VGG-16 Models and CAM Techniques

Byoung Woo Lee¹⁾ · Jae Seok Ryu²⁾ · Se Min Jung³⁾ · Bo Mi
Mun⁴⁾ · Yang Sok Kim⁵⁾

¹²³⁴⁵Management Information Systems, Keimyung University

Abstract

As the number of pet-owning households is increasing, more people care about pet health care. Also, there are a lot of image recognition and classification researches progressing due to the development of machine learning and deep learning. In this paper, we classified skin diseases of dogs by using CNN(convolution neural network) based VGG-16 model which had been modified. As a result, the accuracy of Global Average Pooling was 0.919 and that of Global Max Pooling was 0.926 which shows that the performance of Global Max Pooling was relatively superior. The heat map of region of interest(ROI) highlighted which parts of the image the VGG-16 model was looking at and it did not identify the skin parts of diseases with high accuracy. The reason for this seemed to be the differences in hair' s length and color, and the existences of wrinkles and spots according to the breeds. Especially because skin disease causes hair loss, the model focused on the skin without hair and failed identifying the correct parts of skin diseases. Therefore, for the high accuracy of the model, it requires to train the data divided by the breeds separately and masking the parts of diseased skin previous to classification will lead to classification model with higher accuracy.

Keywords: CNN, Dog' s skin disease, VGG-16, CAM, Classification, ROI

1) Master course student, Management Information Systems, Keimyung University, Daegu 42601, korea

2) Undergraduate student, Management Information Systems, Keimyung University, Daegu 42601, korea

3) Undergraduate student, Management Information Systems, Keimyung University, Daegu 42601, korea

4) Undergraduate student, Management Information Systems, Keimyung University, Daegu 42601, korea

5) Corresponding author: Management Information Systems, Keimyung University, Daegu 42601, korea. E-mail: yangsok.kim@kmu.ac.kr