

HSE Block : SE Block을 활용한 합성곱 신경망 필터 수 자동 최적화

김태욱¹, 정현진², 홍정희^{1*}

¹연세대학교 소프트웨어학부, ²연세대학교 컴퓨터정보통신공학부

HSE Block : Automatic Optimization of the Number of Convolutional Layer Filters using SE Block

Tae-Wook Kim¹, Hyeon-Jin Jung², Ellen J. Hong^{1*}

¹Division of Software, Yonsei University

²Department of Computer & Telecommunications Engineering, Yonsei University

요약 본 논문은 탐색 알고리즘 없이 자동으로 모델의 합성곱 필터의 개수를 최적으로 결정할 방법에 대해 연구하고자 한다. 본 논문은 SENet에서 제안한 SE Block을 합성곱 신경망에 연결하고 하단의 학습하지 않는 합성곱 신경망을 연결한 HSE Block을 제안한다. HSE Block 모델에 두 개의 데이터셋을 이용하여 필터의 개수를 3 epoch 당 1개씩 증가시키는 실험과 필터 내의 값에 따라 필터의 개수를 증가시키는 실험을 수행하였다. 이 실험을 바탕으로 한 층의 HSE Block이 아닌 다층의 HSE Block으로 모델을 구성하고, 기존의 실험할 때 사용한 데이터셋에 비해 더욱 학습하기 어려운 데이터셋을 사용하여 실험을 진행하였다. 기존보다 학습하기 어려운 데이터셋에 대해 HSE Block의 개수를 2개, 3개, 4개, 5개로 두고 실험을 수행함으로써 HSE Block의 효과를 검증하였다.

• **주제어** : 필터, 최적화, SENet, 합성곱 신경망, SE Block' s Excitation

Abstract In this paper, we are going to study how we can automatically determine the number of convolutional filters for the optimal model without a search algorithm. This paper proposes HSE Block by connecting SE Block proposed in SENet to a convolutional neural network and connecting a convolutional neural network not learned at the bottom. An experiment was conducted to increase the number of filters by one per 3 epoch using two datasets for the HSEBlock model and to increase the number of filters by the value in the filter. Based on this experiment, the model was constructed with multi-layer HSE Block instead of layer HSE Block, and the experiment was carried out using a dataset that was more difficult to learn than the one used in the previous experiment. The effect of HSE Block was verified by conducting an experiment with the number of HSE Blocks set to 2, 3, 4, and 5 on a dataset that is more difficult to learn than before.

• **Key Words** : Filter, Optimization, SENet, Convolutional Neural Network, SE Blocks' Excitation

Received 13 September 2022, Revised 27 September 2022, Accepted 29 September 2022

* **Corresponding Author** Ellen. J. Hong, Division of Software, Yonsei University, 1 Yonseidae-gil, Wonju, Gangwon-do, Korea.
E-mail: ellenhong@yonsei.ac.kr

I. 서론

딥러닝 모델의 성능을 최대한으로 끌어올리면서 과적합과 같은 훈련 중에 발생할 수 있는 문제를 방지하기 위해 모델의 하이퍼 파라미터를 적절하게 설정해야 한다. 지금까지의 모델들의 하이퍼 파라미터 설정 방법은 Random Search[1], Grid Search[2] 등이 존재한다. 하지만 이러한 방법들은 다량의 훈련을 반복하면서 하이퍼 파라미터를 찾아야 하므로 상당한 시간을 소모해야 하는 단점이 있다. 이러한 단점을 해결하기 위해 본 논문은 모델의 합성곱 필터의 최적 개수를 자동으로 결정할 수 있는 방법에 대해 연구하고자 한다.

II. HSE Block 검증

2.1 SE Block을 이용한 모델 블록 설계

본 논문은 SENet[3]의 SE Block을 기반으로 Fig. 1과 같은 계층으로 이루어진 HSE Block을 제안하고자 한다. HSE Block은 Fig. 1과 같이 2개의 합성곱 층과 1개의 SE Block으로 구성된다. HSE Block은 SE Block의 상단부와 하단부에 합성곱 층을 연결하였다. 이때, 하단부의 합성곱 층은 학습을 진행하지 않고, Input Shape와 Output Shape가 동일하도록 변형하는 역할을 수행한다. 이러한 구조를 가진 HSE Block을 통해 합성곱 층의 필터 수의 자동 최적화를 기대할 수 있다.

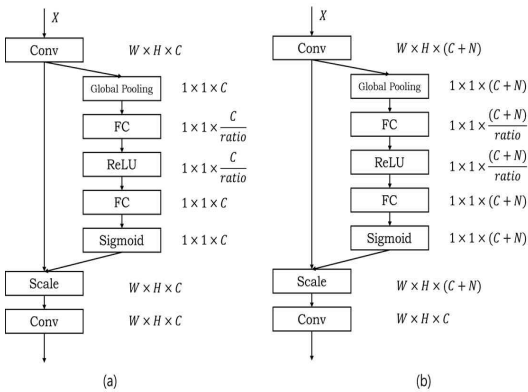


Fig. 1 A HSE Block for Hyperparameter Optimization (a) before filter increasing; (b) after n filter increasing

HSE Block의 상단부 합성곱 층에서는 필터의 개수가 증가할 경우, SE Block ratio 값에 따라 결정되는 새로운 가중치들이 SE Block 내의 FC 2개의 층에 추가되게끔 작동한다. 필터의 개수가 늘어남에 따라 새로운 가중치들로 이루어진 채널이 추가되는 동작은 하단부 합성곱 층에서 동일하게 적용된다. 추가되는 새로운 가중치들은 He-Normal[4] 초깃값을 이용한다. HSE Block에 필요한 하이퍼 파라미터 종류는 SE block의 ratio=r, kernel_size=k, init_filter=i 총 3가지이다. 자세한 블록의 내부 구성은 Table 1과 같다.

Table 1. Layers of HSE Block

Layer	Hyper Parameter
Conv	padding=same, kernel_init=he_normal, kernel_size=k, filter=i, activation=ReLU
GlobalAvgPooling	-
FC	activation=ReLU
FC	activation=sigmoid
Scale	-
Conv	padding=same, kernel_init=he_normal, kernel_size=k, filter=i, activation=ReLU, Train=False

HSE Block 중간에 위치한 SE Block의 Excitation에서 1 epoch의 학습 과정동안 도출되는 모든 벡터에 대하여, 각 원소에 대한 평균을 계산한다. 평균 계산된 행렬의 원소 중 0.9 이상의 값을 가지는 비율을 중요 필터 비율이라고 정의한다. 본 논문에서는 모델의 학습 과정에서 합성곱 층 필터의 개수가 증가할수록 중요 필터의 생성률이 낮아져서 중요 필터 비율은 낮아질 것이라고 가정하였다.

2.2 필터 중요 비율 실험 설계

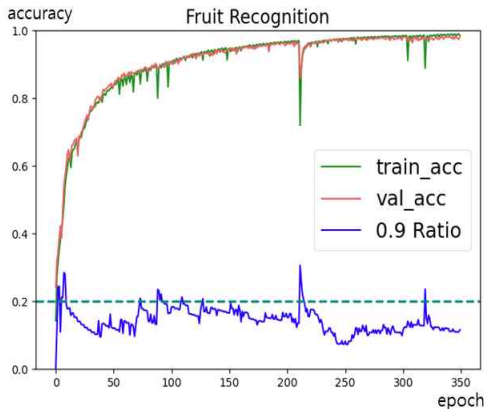
학습 과정에서는 3 epoch마다 상단부 합성곱 층의 필터를 하나씩 증가시키며 중요 필터 비율의 변화를 확인한다. 데이터셋은 Fruit Recognition[5]과 Vegetable Image Dataset[6]을 이용한다. 데이터셋 전처리는 이미지를 224x224 크기로 Resize 하는 방식과 0~1 사이의 값을 갖도록 정규화하는 방식만을 사용한다. 학습 모델은 본 논문에서 제안한 블록, Global Average Pooling, Dense 층 순서로 구성하였다. 각층의 하이퍼 파라미터는 Table 2와 같다.

Table 2. Layers of Single HSE Block Model

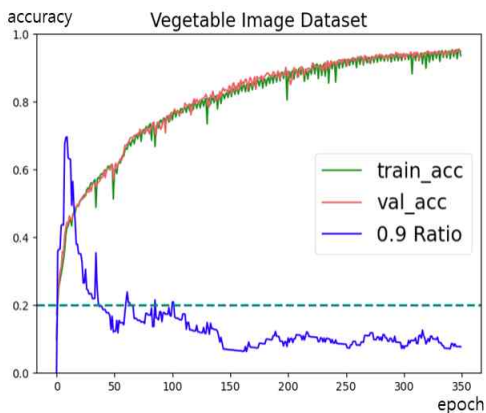
Layer	Hyper Parameter
InputLayer	Input Shape : (224, 224, 3)
HSE Block	ratio=2, init_filter=8, kernel_size=(3, 3)
MaxPooling2D	pool_size=(2, 2), strides=(2, 2)
GlobalAvgPooling	-
Dense	output_node=Number of Class, activation=softmax

2.3 필터 중요 비율 실험 결과

Fig. 2의 (a), (b) 모두 학습 과정에서 합성곱 층 필터 수가 증가할수록 중요 필터 비율이 20% 이하로 하락하는 모습을 볼 수 있다.



(a)



(b)

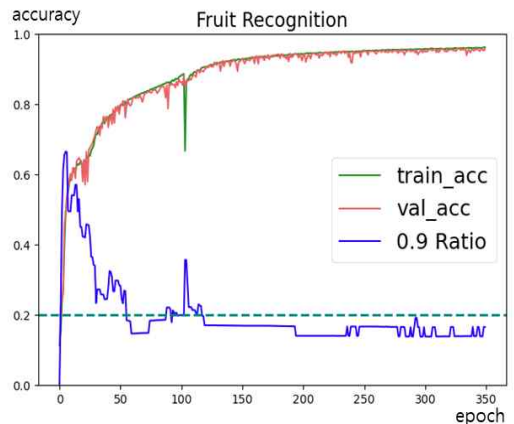
Fig. 2 SE Block Excitation 0.9 Upper Values Ratio: (a) fruit recognition; (b) vegetable image dataset

2.4 필터 수 증가 실험 설계

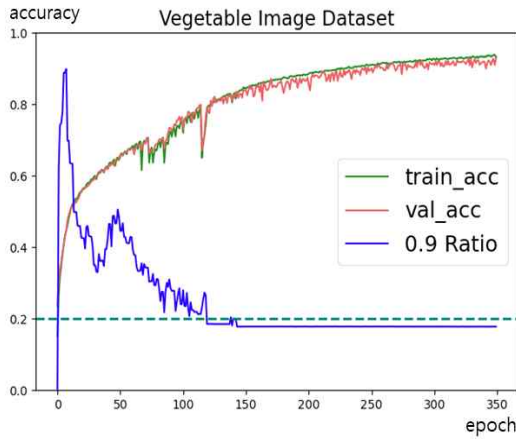
실험 2.3를 통해, 필터를 지속적으로 증가시켰을 때 모델의 정확도가 증가했으며, 중요 필터 비율이 0.2 이하로 하락함을 알 수 있다. 이는 두 개의 데이터셋을 통해 생성할 수 있는 중요한 필터 수가 제한적임을 나타낸다고 볼 수 있다. 이를 바탕으로 중요 필터 비율 0.2를 필터 증가 여부의 Threshold 값으로 설정하였다. 이후, Threshold 값에 따라 필터를 자동으로 증가시켜 모델이 적절한 필터 수를 찾아가며 학습하는지 실험하였다. 다만, 상단부 합성곱 층 초기 필터 수가 8개일 때, 학습 초기에 드물게 중요 필터 비율이 20%를 넘지 못하여 학습이 원활하게 진행되지 않는 경우가 발생하였다. 이는 초기 필터 수가 너무 적어 중요 필터를 찾는 과정을 포함한 전반적인 학습 자체가 원활하게 진행되지 않았다고 해석할 수 있다. 이를 해결하기 위해 epoch 50까지 3 epoch마다 합성곱 층 필터 수를 증가시키는 방식을 사용하여 실험을 진행하였다.

2.5 필터 수 증가 실험 결과

Table 3은 실험 결과를 요약한 표이다. 필터 수는 실험 2.3에 비해 Fig. 3의 (a)는 3.65배, (b)는 2.53배 감소하였고, 학습 시간은 각각 약 1.7배, 1.54배 감소하였다. 그러나 정확도는 실험 2.3에 비해 Fig. 3의 (a)는 0.024, (b)는 0.021 하락하였다.



(a)



(b)

Fig. 3 SE Block Excitation 0.9 Upper Values Ratio: (a) fruit recognition; (b) vegetable image dataset

Table 3. Comparison Result

필터증가조건	Fruit Recognition		Vegetable Image Dataset	
	계속 증가	Proposed	계속 증가	Proposed
정확도	0.982	0.958	0.949	0.928
필터 수(개)	124	34	124	49
학습시간(초)	48614.24	28604.36	22299.26	14522.14

III. 다층 HSE Block 모델 검증

3.1 다층 HSE Block 모델 실험 설계

실험 2는 본 논문에서 제안한 HSE Block이 올바르게 필터 수를 자동으로 최적화하는지 검증하기 위해 하나의 블록을 사용하였으며, 하나의 블록만으로 높은 정확도에 도달할 수 있는 데이터셋을 사용하였다. 따라서, 이 실험에서는 해당 블록을 여러 층으로 쌓았을 경우, 실험2에서 사용한 데이터셋보다 학습하기 어려운 데이터셋에서도 각 블록이 효과적으로 필터 수를 자동으로 최적화를 수행하는가에 대한 검증을 진행한다. Table 4와 같이, HSE Block을 2, 3, 4, 5개씩 MaxPooling 층과 함께 쌓아 중요 필터 비율과 정확도를 확인한다.

Table 4. Layers of Multiple HSE Block Model

Layer	Hyper Parameter
InputLayer	Input Shape : (224, 224, 3)
HSE Block	ratio=2, init_filter=8, kernel_size=(3, 3)
MaxPooling2D	pool_size=(2, 2), strides=(2, 2)
:	:
HSE Block	ratio=2, init_filter=8, kernel_size=(3, 3)
MaxPooling2D	pool_size=(2, 2), strides=(2, 2)
GlobalAvgPooling	-
Dense	output_node=Number of Class, activation=softmax

실험 2에서 필터 수가 적을 경우, 중요 필터 비율이 0.2 이상으로 상승하지 못하고 필터가 더 이상 증가하지 못함으로써 학습이 원활하게 진행되지 않는 경우가 드물게 발생하였다. 이를 해결하기 위해 층이 깊어짐으로 인해 필터의 수가 크게 증가할 것을 고려하여, 필터를 학습 진행과정에서 epoch 50까지는 3 epoch 주기로 합성곱 층 필터 수를 증가시키는 방식이 아닌 초기 필터 수를 16으로 설정하여 실험을 진행하였다.

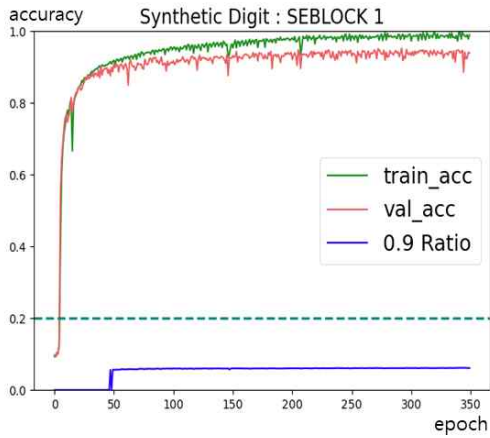
학습 과정에서는 3 epoch마다 각 HSE Block의 중요 필터 비율 값이 0.2를 초과하는 경우 필터를 증가시킨다. 데이터셋은 Synthetic Digits[7]을 이용한다. 데이터셋 전처리는 이미지를 224x224 크기로 Resize 하는 방식과 0~1 사이의 값을 갖도록 정규화하는 방식만을 사용한다.

3.2 다층 HSE Block 모델 및 실험 결과 분석

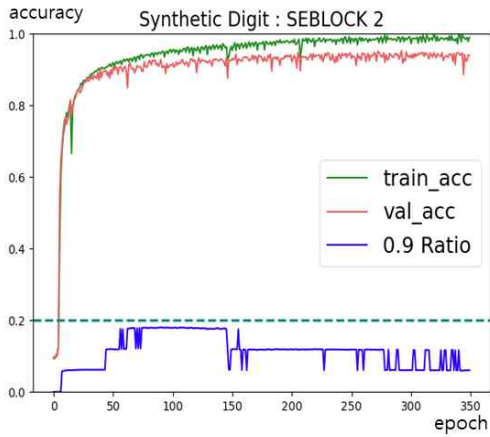
Table 5는 실험 결과를 요약한 표이다. Fig. 4 (a)-(e)는 각각 3.1의 실험 중 5개의 HSE Block으로 구성된 모델의 학습 결과이다. 정확도는 5개의 HSE Block으로 구성된 모델이 0.939로 가장 높으며, 필터 수는 총 256개이다. 이때, Fig. 4 (e)의 중요 필터 비율은 다른 SE Block에 비해 필터가 늘어남에도 0.2를 초과하였다. 이는 3번째 HSE Block에서 도출되는 Feature map에 추출할 수 있는 중요한 특성이 많아 높은 정확도임에도 불구하고 많은 필터를 지속적으로 생성하는 것으로 해석할 수 있다.

Table 5. Comparison Result

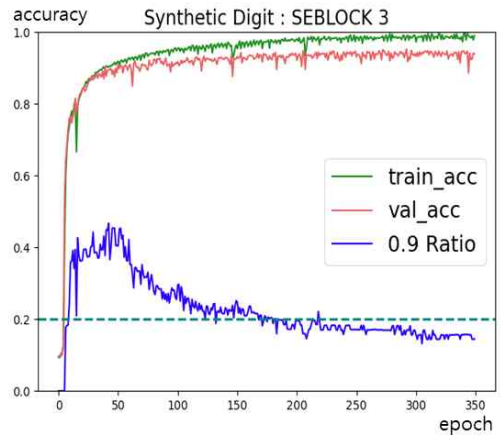
HSE	상단 합성곱 층 필터 수						정확도
	1	2	3	4	5	Total	
2	16	126	-	-	-	142	0.572
3	17	82	130	-	-	229	0.636
4	16	25	28	130	-	199	0.871
5	16	16	75	131	18	256	0.939



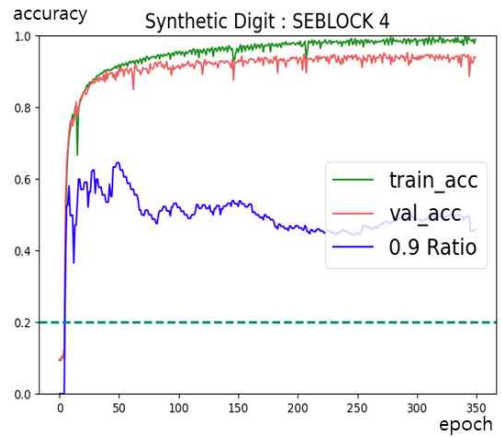
(a)



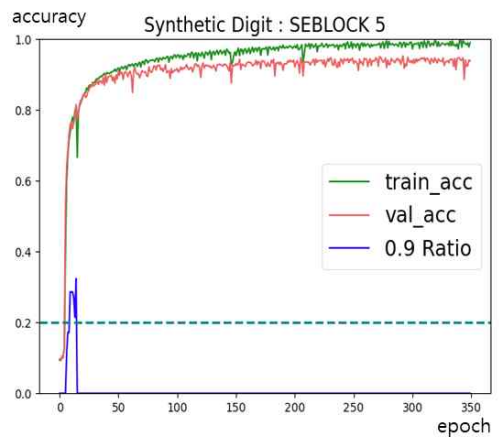
(b)



(c)



(d)



(e)

Fig. 4 SE Block Excitation 0.9 Upper Values Ratio:

(a) SE Block 1; (b) SE Block 2; (c) SE Block 3;

(d) SE Block 4; (e) SE Block 5;

IV. 결론 및 향후 연구

본 논문은 학습에 많은 시간이 소요되는 상황에서 하이퍼 파라미터 탐색 알고리즘 없이 자동으로 최적의 필터 개수를 빠르게 찾을 수 있는 방법을 제안하였다. 또한 필터 수를 증가시키기 위해 기존에 지표로 사용되었던 정확도가 아닌 SE Block의 Excitation 과정에서 도출된 값을 새로운 지표로 제시하였다. 기존의 SE Block이 다른 합성곱 층에 간편하게 적용할 수 있다는 장점을 유지함으로써 제안한 블록 또한 다른 합성곱을 간편하게 대체하여 적용할 수 있다. 하지만, 실험 3과 같이 필터가 정확도에 비해 필요 이상으로 생성될 수 있다는 단점이 있다. 또한 3개의 데이터셋에 대해서만 실험을 진행하였기 때문에 일반화를 위해서는 더 다양한 데이터셋을 활용하여 결과를 분석할 필요가 있다. 더불어, SE Block의 ratio 값을 초기 필터가 작음을 고려하여 2로 설정하였으나 초기 필터 수를 16 이상으로 설정하여 SE Block의 ratio 값을 늘려가며 적절한 값을 찾아야 할 필요가 있다. 따라서, 적합한 SE Block의 ratio 값을 찾고, 최대 필터 수를 제한하여 필터 수를 자동으로 최적화하는지에 대한 검증은 추후 연구로 진행하고자 한다.

ACKNOWLEDGMENTS

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2022R1F1A1074273).

REFERENCES

[1] Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2).

[2] Liashchynskiy, P., & Liashchynskiy, P. (2019). Grid search, random search, genetic algorithm: A big comparison for NAS. *arXiv preprint arXiv:1912.06059*.

[3] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7132-7141).

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034).

[5] Mureşan, H., & Oltean, M. (2017). Fruit recognition from images using deep learning. *arXiv preprint arXiv:1712.00580*.

[6] Li, Z., Li, F., Zhu, L., & Yue, J. (2020). Vegetable recognition and classification based on improved VGG deep learning network model. *International Journal of Computational Intelligence Systems*, 13(1), 559-564.

[7] Roy, P., Ghosh, S., Bhattacharya, S., & Pal, U. (2018). Effects of degradations on deep neural network architectures. *arXiv preprint arXiv:1807.10108*.

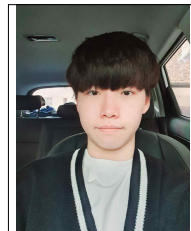
저자소개

김 태 옥 (Tae-Wook Kim)



2021년 3월~현재 : 연세대학교
소프트웨어학부(재학)
관심분야 : 딥러닝, 머신러닝,
컴퓨터비전

정 현 진 (Hyeon-Jin Jung)



2020년 3월~현재 : 연세대학교
컴퓨터정보통신공학부(재학)
관심분야 : 딥러닝, 머신러닝,
컴퓨터비전

홍 정 희 (Ellen J. Hong)



2023년 2월 : KAIST
전기및전자공학과(공학박사)
2016년 10월~2018년 1월 :
동서대학교 컴퓨터공학부 조교수
2018년 2월~2019년 8월 : KT
융합기술원 선임연구원
2019년 9월~현재 : 연세대학교 소프트웨어학부 조교수
관심분야 : 인공지능, 시스템 모델링 시뮬레이션,
시뮬레이션 기반 최적화, 디지털 트윈, 지능형 시스템