

# 합성곱 신경망의 마스크 기반 선택적 다운샘플링 기법

곽 철 영\*, 박 세 웅<sup>o</sup>

## Mask-Based Selective Downsampling in Convolutional Neural Networks

Chulyoung Kwak\*, Saewoong Bahk<sup>o</sup>

### 요 약

본 논문에서는 합성곱 신경망(Convolutional neural network)의 효율적인 동작을 위한 선택적 다운 샘플링 기법에 관한 연구를 진행한다. 합성곱 신경망의 피쳐 레이어의 중요도에 따라 해상도를 달리 사용함으로써, 계산량을 줄일 수 있는 선택적 다운 샘플링 기법에 대해 살펴본다. 비교 기법보다 80% 경량화된 네트워크를 설계하였고, 경량화된 네트워크의 성능 하락을 완화시키기 위한 방법론을 제안하였고, 실험을 통해 성능을 평가하였다.

**Key Words** : Down sampling, Convolutional Neural network, Signal Processing

### ABSTRACT

This letter presents a study on selective down-sampling techniques designed to enhance the efficiency of convolutional neural networks (CNNs). By selectively adjusting the resolution of feature layers based on the down-sampling mask, this method aims to improve efficiency. We develop a network that is 80% lighter than the baseline scheme, and introduce several methods to mitigate the performance degradation of light weight networks. The performance of the proposed methods is evaluated through extensive experiments.

## I. 서 론

최근 딥러닝 기술의 발달로 딥러닝 기반 어플리케이션이 증가하고 있다. 자원이 풍부한 서버와는 달리 실제 어플리케이션이 사용되는 모바일, 엣지 디바이스, 모빌리티 등에서는 제한된 자원으로 딥러닝 어플리케이션을 실행해야 하기 때문에 양자화, 프루닝 등의 효율적인 추론을 위한 연구가 많이 이루어 지고 있다<sup>[1]</sup>. 다운 샘플링 기법은 신호처리 분야에서 널리 쓰이는 기법으로, 입력 신호를 축소하거나 압축하여 해상도를 줄여 효율적인 연산을 할 때 사용되는 기법이다.

합성곱 신경망의 경우 이미지 속의 지역적인 정보를 바탕으로, 객체의 분류(classification), 감지(detection), 세그멘테이션(segmentation) 등 여러가지 목적에 사용된다. 합성곱 신경망은 입력 이미지의 지역적인 특징을 파악하는데 사용되며, 모든 정보를 동등하게 간주하지 않는다. 합성곱 신경망의 피쳐 레이어의 또한 비슷한 방식으로 모든 정보가 동일하게 중요하지 않을 수 있다.

본 연구에서는 합성곱 신경망의 피쳐 레이어의 선택적 다운 샘플링을 위한 개선된 기법을 제안하며, 실험 통해 제안된 다운 샘플링 기법의 성능을 평가한다.

## II. 시스템 모델

### 2.1 마스크 기반 선택적인 다운 샘플링

일반적으로 다운 샘플링은  $H \times W \times C$  크기의 입력 신호를  $H/d \times W/d \times C$ 로 축소시켜주는 기법이다. 합성곱 신경망의 경우 공간적 해상도를 조절할 때, 풀링(Pooling) 혹은 stride를 조절한 합성곱 레이어를 통해 다운 샘플링을 수행 한다.

마스크 기반 선택적 다운 샘플링<sup>[4]</sup>의 경우 입력에 대해  $(H/d) \times (W/d)$  크기의 마스크  $M$ 을 사용해, 마스크의 값에 따라 마스크의 값이 0인 경우 기존 해상도로, 마스크의 값이 1인 경우 다운 샘플링된 값을 사용한다.

$$y(x, M) = (f(x_1, M), \dots, f\left(\frac{x_H \times W}{d \times d}, M\right))$$

전체 피쳐맵을 가로, 세로 각각  $d$ 개의 패치로 나뉘었을 때,  $k$ 번째 패치에 포함된  $x_{k,j}$ 에 대해

\* First Author : Hyundai Motors Company, cykwak1218@gmail.com, 정회원

<sup>o</sup> Corresponding Author : Seoul National University Department of ECE and INMC, sbahk@snu.ac.kr, 정교수, 종신회원  
논문번호 : 202412-310-A-LU, Received December 3, 2024; Revised December 23, 2024; Accepted December 23, 2024

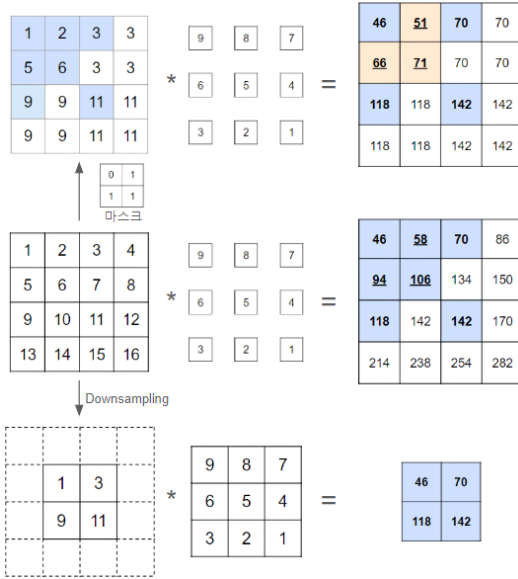


그림 1. 다운 샘플링 기법 동작 비교 (위: 선택적 다운 샘플링, 가운데: 확장 합성곱, 아래: 간격 2인 다운 샘플링된 컨볼루션)

Fig. 1. Comparison of Down-sampling Methods (Top: Mask-based Selective Down-sampling, Middle: Dilated Convolution, Bottom: Standard Convolution with stride 2)

$$y_{i,j} = f(x_k, M) = \begin{cases} x_{i,j} & \text{if } m_{i,j} = 0 \\ \text{Pool}(x_{i,j}) & \text{if } m_{i,j} = 1 \end{cases}$$

과 같이 표현할 수 있다. 그림 1과 같이 합성곱을 연산하기 전 입력 신호에 마스크를 적용해준 뒤 희소(sparse) 합성곱<sup>[3]</sup> 방법을 이용하여, 저해상도 영역의 반복된 계산을 줄일 수 있다. 최종적으로 생성된 피쳐맵의 경우 확장 합성곱과 동일한 공간 크기를 가져 표현력은 유지하면서도, 계산량은 다운 샘플링의 활성화도에 따라 줄어들게 된다.

## 2.2 네트워크 디자인

[4]에서 제안된 기법의 경우, 마스크 예측을 위한 모델이 일반적인 합성곱 신경망으로 구성되어 있다. 선택적 다운 샘플링의 경우 특성상 표현력을 유지하면서도 연산량을 줄이기 위해서는 큰 공간적 해상도를 가지는 레이어에 적용해야 하며, 합성곱 신경망의 경우 공간적 해상도에 따라 계산량이 늘어나게 된다. 계산을 효율적으로 하기 위해 제안된 기법인 만큼, 연산량을 줄이기 위해 깊이별 분리(Depthwise-separable) 합성곱 신경망으로 마스크 예측 모델의 구조를 변경하였고, 기존 구조대비 80%의 계산량을 줄일 수 있었다.

또한 선택적 다운 샘플링을 하게 되면, 다운 샘플링

마스크가 적용된 부분의 경우(그림 1의 푸른색 영역의 볼드 표시된 값) 기존의 다운 샘플링을 통한 값과 동일한 값을 얻을 수 있으나, 저해상도 계산을 하기 위한 영역의 값이 변경되기 때문에 고해상도 영역의 값(그림1의 붉은 영역의 밑줄 표시된 값)이 변경되게 된다. 마스크 예측 모델을 경량화 하면서 표현력이 줄어들기 때문에, 이러한 문제가 더 심각해질 수 있다. 따라서, 일정한 위치의 값으로 다운 샘플링을 하지 않고 풀링 기법을 사용하여 테스트의 손실함수를 바탕으로 성능 하락을 줄일 수 있게 하였다.

## 2.3 손실 함수

제한한 예측 모델의 경우, 다운 샘플링이 많이 실행될 경우 성능 하락이 발생할 수 있으며, 적게 실행될 경우 효과가 없게 된다. 따라서, 적정수준의 다운 샘플링이 필요하며, 이를 위해 마스크의 희소성(sparsity)을 손실 함수로 사용하였다. 희소성과 세그멘테이션의 성능의 균형을 위해 세그멘테이션의 손실 함수와 희소성을 합친 값을 최종 손실 함수로 사용하였다. 희소성을 전체 마스크의 크기 대비 다운 샘플링의 비율인

$$S = \frac{\sum M}{H \times W}$$

로 정의하였을 때 최종 손실함수는

$$L = \text{Loss}_{\text{seg}} + \beta * S(1 - S)$$

를 사용하였다.

## III. 실험 결과

성능 평가를 위해 ResNet101<sup>[6]</sup>을 백본 네트워크로 하는 DeepLab V3<sup>[1]</sup> 모델을 사용하였고, 19개의 클래스를 가지는 의미론적 세그멘테이션(Semantic segmentation)을 목적으로 하는 Cityscape<sup>[5]</sup> 데이터셋을 활용하였다. 학습과 추론에는 Nvidia A40 GPU를 사용하였고, 하이퍼 파라미터는 DeepLab V3의 학습 파라미터와 동일하게 사용하였다.

실험 결과는 표 1과 같다. 기존의 기법 대비 깊이별 분리 합성곱 신경망을 사용하여 마스크 예측 모델만 간소화 할 경우(Prop. 1) 마스크 연산을 위한 계산량은 80% 이상 간소화 시킬 수 있으나, 세그멘테이션 성능(mIoU)이 소폭 하락하는 것을 확인할 수 있다. 이는 앞에서 설명한 바와 같이, 마스크 네트워크의 경량화로

표 1. Cityscape 데이터셋에 대한 비교 기법 및 제안 기법의 segmentation 결과와 추가적인 FLOPS 비교.

Table 1. Performance comparison of the proposed method and comparative methods on the Cityscapes dataset.

|             | Light Mask | Adaptive Pooling | Loss | mIoU | FLOPS |
|-------------|------------|------------------|------|------|-------|
| Baseline[4] | -          | -                | -    | 0.77 | 690M  |
| Prop. 1     | O          | -                | -    | 0.74 | 130M  |
| Prop. 2     | O          | O                | O    | 0.77 | 140M  |

인한 성능 하락과 더불어 선택적 다운 샘플링 계산에서 발생하는 고해상도 영역의 왜곡으로 인한 것으로 생각된다. 이를 방지하기 위한 선택적 풀링 기법을 적용하고 로스 함수를 변경하면(Prop. 2), 풀링에 사용되는 계산량으로 인해 계산량이 소폭 증가하지만, 비교 기법과 동일한 수준의 세그멘테이션 성능을 확인할 수 있었다.

손실 함수의 경우 해당 테스트에 대해 유의미한 성능 차이를 보이지 않았으나, 기존의 방법론의 희소성의 경우 타겟 희소비율을 설정해야 하기 때문에, 정확한 값을 알기 어려워 학습을 방해하는 문제가 있을 수 있다. 따라서, 제안하는 손실 함수를 사용하면 안정적인 학습 효과가 있을 것으로 예상된다.

#### IV. 결 론

본 논문에서는 합성곱 신경망의 효율적인 동작을 위한 선택적 다운 샘플링 기법에 관한 연구를 진행하였다. 합성곱 신경망의 피쳐 레이어의 공간적 해상도를 선택적으로 다운 샘플링 함으로써, 성능 하락 없이 계산량을 줄일 수 있는 선택적 다운 샘플링 기법에 관한 연구를 진행하였다. 비교 기법 대비 80% 경량화 된 네트워크를 설계하였고, 경량화로 인한 성능 하락을 완화 시키기 위한 방법을 제안하였고, 실험을 통해 제안 기법의 성능을 확인할 수 있었다.

#### References

- [1] L.-C. Chen, et al., "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. ECCV*, 2018.
- [2] M. J. A. Shanto, et al., "Lightweighted real-time object detection on a custom edge device," *J. KICS*, vol. 49, no. 10, pp. 1447-1457, Oct. 2024.
- [3] G. Benjamin and D. M. Laurens, "Submanifold sparse convolutional networks," *arXiv preprint arXiv:1706.01307*, 2017.
- [4] R. Hesse, et al., "Content-adaptive down-sampling in convolutional neural networks" in *Proc. IEEE CVPRW*, 2023.
- [5] M. Cordts, et al., "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE CVPR*, 2016.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, 2016.