Spectral Pooling: DFT 기반 풀링 계층이 보여주는 여러 가능성에 대한 연구

이성주 조남익

서울대학교 전기정보공학부 뉴미디어통신공동연구소

thomas11809@snu.ac.kr nicho@snu.ac.kr

Spectral Pooling: A study on the various possibilities of the DFT-based Pooling layer

Lee, Sung Ju Cho, Nam Ik

Department of ECE, INMC, Seoul National University

요약

GPU의 발전과 함께 성장한 딥러닝(Deep Learning)은 영상 분류 문제에서 최고의 성능을 보이고 있다. 그러나 합성곱 신경망 기반의 모델을 깊게 쌓음에 따라 신경망의 표현력이 좋아짐과 동시에 때로는 학습이 잘되지 않고 성능이 저하되는 등의 부작용도 등장했다. 성능 향상을 방해하는 주요 요인 중 하나는, 차원감소 목적에 따라 필연적으로 정보 손실을 겪어야 하는 풀링 계층에 있다. 따라서 특성맵(Feature map)의 차원감소를 통해 얻게 되는 비용적 이득과 모델의 분류 성능 사이의 취사선택(Trade-off)이 존재한다. 그리고 이로부터 자유로워지기 위한 다양한 연구와 기법이 존재하는데 Spectral Pooling도 이 중 하나이다. 본 논문에서는 이산 푸리에 변환(Discrete Fourier Transform, DFT)을 이용한 Spectral Pooling에 대한 소개와, 해당 풀링의 성질을 통상적으로 사용되고 있는 Max Pooling과의 성능 비교를 통해 분석한다. 또한 영상 내 고주파수 부분에서 특히 더 강건하지 못하다는 맥스 풀링의 고질적인 문제점을, Spectral Pooling과의 하이브리드(Hybrid) 구조를 통해 어떻게 극복해나갈 것인지 그 가능성을 중심으로 실험을 수행했다.

1. 서론

영상 분류 문제(Image Classification task)는 영상처리 분야의 주된 연구 주제 중 하나이다. 이 분야에서 선형대수학과 통계학적 지식을 기반으로 한 여러 기계 학습 기법들이 과거의 주류를 이끌었던 반면, 근 10년 사이에는 GPU 하드웨어의 발전에 힘입어 딥러닝(Deep Learning)[1]을 이용한 영상 분류 기법 연구들이 활발히 이루어지고 있다. 특히, GPU 병렬 연산의 이점을 많이 활용하는 합성곱 신경망(Convolutional Neural Networks, CNNs)[2]을 기반으로 한 모델들이 최고 수준(state-of-the-art, SOTA)의 성능을 보여주게 되면서, 해당 신경망이 여타 학습 모델들과 차별성을 가지는 이유에 대한 수 많은 연구도 함께 수행되었다.

합성곱 신경망을 구성하는 요소들은 일반적으로 다음과 같은 계층들의 묶음으로 이루어져 있다.

- 합성곱 계층(Convolution layer)
- 활성 함수 (Activation function)
- 풀링 계층 (Pooling layer)

여기서 풀링 계층은 앞선 계층의 특성맵(Feature map)을 입력으로 받아 앞단보다 작은 크기(height, width)의 특성맵으로 출력하는 역할을 한다. 영상의 크기를 부표본화(Subsampling)하여 입력 데이터의 차원 감소를 통해 학습에 필요한 비용을 줄이는 것이 풀링 계층을 사용하는 이유이다.

딥러닝 모델은 신경망이 깊어지면 깊어질수록 미세한 특징들(Features)의 역할이 더 중요해진다. 풀링 계층을 사용하면 입력 정보의 손실을 피할 수 없게 되고, 이는 곧 특성맵에 나타나는 세밀한 특징들을 그대로 버리게 되는 것이므로 분류 성능 저하를 일으킬 수 있다. 따라서 차원감소를 통해 얻게 되는 비용적 이득과, 모델의 분류 성능 사이의 취사선택(Trade-off)에서 자유로워지기 위한 다양한 연구가 계속되어왔다. 통상적으로 사용되는 맥스 풀링 계층(Max Pooling layer) 이외에도 여러 종류의 방법들이 연구되었고[3], 풀링 계층을 사용하지 않고 합성곱 계층의 보폭(stride)을 늘려 차원감소를 꾀하는 방법[4]도 제안되었다.

본 논문에서는 이산 푸리에 변환(Discrete Fourier Transform, DFT)을 이용한 풀링인 Spectral Pooling[5]에 대한 소개와 함께, 통상적으로 자주 사용되는 Max Pooling과의 성능 비교 및 분석을 한다. 또한 두 풀링 간의 여러 조합을 구성하여 얻은 결과를 바탕으로, 서로의 장단점이 상호보완적으로 작용할 수 있다는 가능성을 보인다.

2. 관련 연구

2.1. 풀링 계층 (Pooling layer)

합성곱 신경망에서 풀링 계층은 주로 합성곱 계층 또는 활성 함수의 뒤에 위치한다. 풀링은 입력으로 받은 특성맵의 크기를 줄여주는데, 이는 불필요한 정보를 버리고 손실(Loss)에 의해 정의된 전역 특징(Global Features)을 남기는 역할을 한다.

그림 1. Max Pooling과 Spectral Pooling을 비교한 사진: 차원을 급격히 감소시키면(Sharp Reduction), 맥스 풀링의 경우 고양이의 눈과 귀 등의 특징들을 잘 찾기 어려운 반면, Spectral Pooling의 경우 저주파수대의 정보를 많이 보존하고 있어 특징을 찾기에 비교적 수월함을 확인할 수 있다.

풀링 계층을 통해 얻을 수 있는 비용적인 이득은 크게 세 가지로 생각할 수 있다. 먼저 입력 데이터의 감소된 차원에 상응하여 저장소(memory) 요구량이 감소한다. 또한, 모델 파라미터의 수가 줄어들기 때문에 학습과 평가에 필요한 전체 연산량(computational cost)도 감소하여 학습 속도가 증가한다. 마지막으로 적당히 얕은 모델의 경우 파라미터의 수가 적다는 것은 신경망의 표현력(expressive power)이 작다는 것을 의미하고, 이는 곧 학습 데이터에 대한 과적합(Overfitting) 문제를 방지하는 데 도움이 된다.

2.2. 맥스 풀링 계층 (Max Pooling layer)

맥스 풀링은 특성맵의 국부 수용영역(local receptive field)에서 최댓값을 취하여 차원을 감소시키는 풀링 방법이다.

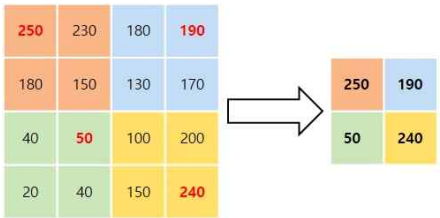

그림 2. Max Pooling 알고리즘 예시

맥스 풀링의 문제점은, 영상 내 고주파수 부분에서 다른 풀링들보다 특히 더 강건하지 못하다는 것인데, 그 이유에 해당하는 예는 다음과 같다[3]: 물체의 가장자리(edge)에 해당하는 픽셀값은 작지만 주변의 값은 클 때 풀링에 의한 출력 특성맵은 주변의 값으로 고정되게 된다. 따라서 큼직한 특징들은 비교적 잘 구분하는 반면, 직선이나 털과 같은 세밀한 특징들은 정보 손실로 놓치게 된다. 또한, 영상 내 잡음(noise)이 많은 경우에도 취약한 모습을 보인다. 이러한 문제점은 딥러닝 모델의 신경망이 깊어질수록 더 나쁜 결과를 낳게 된다. **그림1.**을 보면 차원이 많이 감소할수록 세밀한 부분의 특징을 더욱 찾을 수 없음을 확인할 수 있다.

2.3. 이산 푸리에 변환 (Discrete Fourier Transform, DFT)

이산 푸리에 변환(DFT)은 시공간의 신호를 (spatiotemporal signal) 주파수 영역(Frequency domain)으로 변환하는 방법이다. DFT는 복소 지수함수(Complex exponential function)에 내적하여 얻은 주파수 성분을 통해, 신호의 주파수 분석에 주로 사용된다.

DFT와 그 역변환(Inverse DFT, IDFT)은 행렬 곱으로 나타낼 수 있는 선형변환이다. 특히, 통상적으로 사용되는 2D DFT의 크기를 정규화한 식은 다음과 같다.

$$DFT(X)_{h,w} = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_{m,n} \exp\left(-2\pi i \left(\frac{mh}{M} + \frac{nw}{N}\right)\right)$$

이처럼 크기를 정규화하면 DFT를 나타내는 행렬은 유니터리 행렬(Unitary matrix)이 되고, 이 행렬의 켤레 복소수 전치(Hermitian conjugate)가 곧 역변환을 나타내는 행렬이 된다. 한편, DFT는 실제 구현에서 위 공식으로 직접 만들어지지는 않고, 대신 고속 푸리에 변환(Fast Fourier Transform, FFT)을 통해서 구현된다. DFT의 시간 복잡도(time complexity)는 $O(N^2)$인데 반해, FFT의 경우 $O(N \times \log N)$이다.

3. Spectral Pooling

특성맵의 차원감소를 위해 *O Rippel et al* 이 제안한 Spectral Pooling[5]의 중심 개념은, 주파수 변환을 이용한 저주파 통과 필터(Low-Pass filter, LPF)라고 말할 수 있다. 주파수 영역에서 이상적인 주파수 절단(cropping)을 통해 입력의 크기를 줄이고, 이후 역변환을 통해 원래의 공간 영역(Spatial domain)으로 가져오는 방식이다. 저자가 저주파 통과 필터를 사용한 이유는, 자연적인 영상(natural image)의 전력(power) 기댓값이 통계적으로 저주파 영역에 대부분 밀집되어 있다는 Inverse Power Law를 따르기 때문이다[5, 6]. 따라서 차원을 감소

시켰음에도 불구하고, 입력에서 미미한 정보량만 잃어버린 채 출력을 얻을 수 있다.

3.1. Spectral Pooling 알고리즘

Spectral Pooling 방법의 첫 단계에서는 특성맵을 FFT 변환하고, 그 결과에서 원하는 크기의 저주파수 영역을 절단한다. 그리고 출력을 복소수가 아닌 실수로 만들기 위해 공간 영역으로 변환하기 이전, 켤레 대칭 중심점들의 허수부를 제거한다. 마지막으로 역변환(Inverse FFT, IFFT)을 통해 출력을 얻는다.

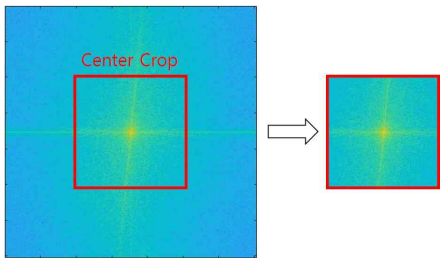

그림 3. Spectral Pooling 알고리즘 중심 개념

해당 알고리즘의 장점은 다른 풀링들에 비해 유연하게 차원을 감소시킬 수 있다는 것이다[5]. 또한 별도의 파라미터를 요구하지 않아서, 무시할만한 연산량만 조금 늘어나는 것 외에는 저장소에 부담을 주지 않는다는 것이다. 그리고 앞서 말한 바와 같이 정보량을 최대한 보존할 수 있기 때문에, 모델이 학습 데이터를 효율적으로 사용할 수 있게 한다.

4. 실험 설계 및 결과

4.1. 실험 모델 구조 (Model Architecture)

본 실험은 Spectral Pooling과 Max Pooling의 성능을 비교하고 분석하는 것에 목적을 두고 있다. 따라서 영상 분류 문제를 해결하는 SOTA 모델을 사용하는 것보다, 경량화된 간단한 모델 구조를 선택하는 것이 합리적이라고 판단했다. 또한, 학습 초반의 손실 수렴양상 및 성능을 분석하는 것에 초점을 맞추어 초매개변수(hyperparameter)를 설계했다.

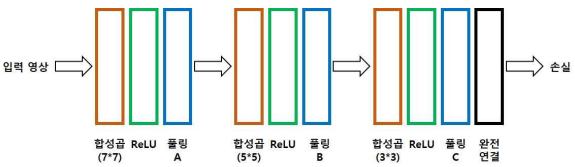

그림 4. 실험 모델 구조

표 1. 실험 모델 초매개변수 세부사항

Dataset	CIFAR-10 / CIFAR-100
Batch size	64
Optimizer	Adam
Train epochs	10
Learning rate	0.0032
Weight decay rate	0.001

4.2. 실험 결과

4.2.1. 실험 1: 학습 데이터의 레이블에 따른 비교

해당 실험에서는 학습 데이터 CIFAR-10과 CIFAR-100에 대한 결과를 비교했다. **그림4**의 모델 구조에서, 풀링 A/B/C가 모두 Max Pooling이면 'Max'로 표기하고, 모두 Spectral Pooling인 경우엔 'Spectral'로 표기하였다.

표 2. [실험1] 결과: 학습 데이터와 상관없이 Spectral Pooling의 성능(빨강 글씨)이 더 좋다. Spectral Pooling은 Max Pooling보다 레이블이 많은 데이터에 강세를 보인다.

Test Accuracy	CIFAR-10	CIFAR-100
Max Pooling	0.6632	0.2806
Spectral Pooling	0.6704	0.3124

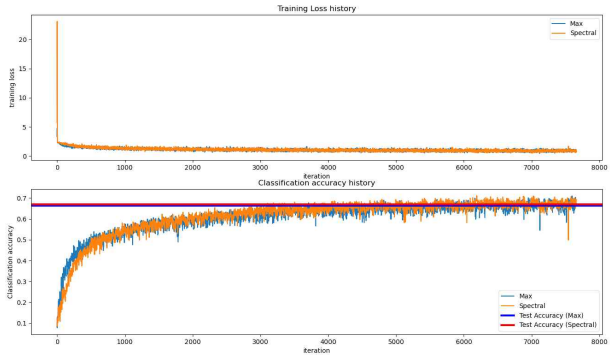

그림 5.(a) [실험1] CIFAR-10: CIFAR-10 학습 데이터에 대한 Max Pooling과 Spectral Pooling의 손실 및 정확도 그래프.

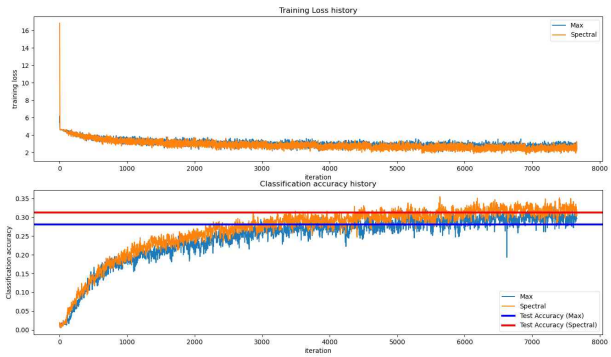

그림 5.(b) [실험1] CIFAR-100: CIFAR-100 학습 데이터에 대한 Max Pooling과 Spectral Pooling의 손실 및 정확도 그래프.

표 2.와 그림 5.(a), (b)를 보면 두 경우 모두 Spectral Pooling의 테스트 정확도(test accuracy)가 Max Pooling보다 높은 것을 확인할 수 있다. 또한 레이블의 수가 많을수록 난이도가 높은 영상 분류 문제이므로 CIFAR-10(레이블 수: 10)에 비해 CIFAR-100(레이블 수: 100)의 학습이 아직 덜 이루어진 것을 볼 수 있다. 따라서 CIFAR-100으로 한 실험이 학습 초반을 더 잘 나타내고 있고, **그림5.(b)**에서 보이듯 테스트 정확도의 격차가 CIFAR-10(**그림.5.(a)**)의 격차보다 크게 벌어져 있음을 알 수 있다.

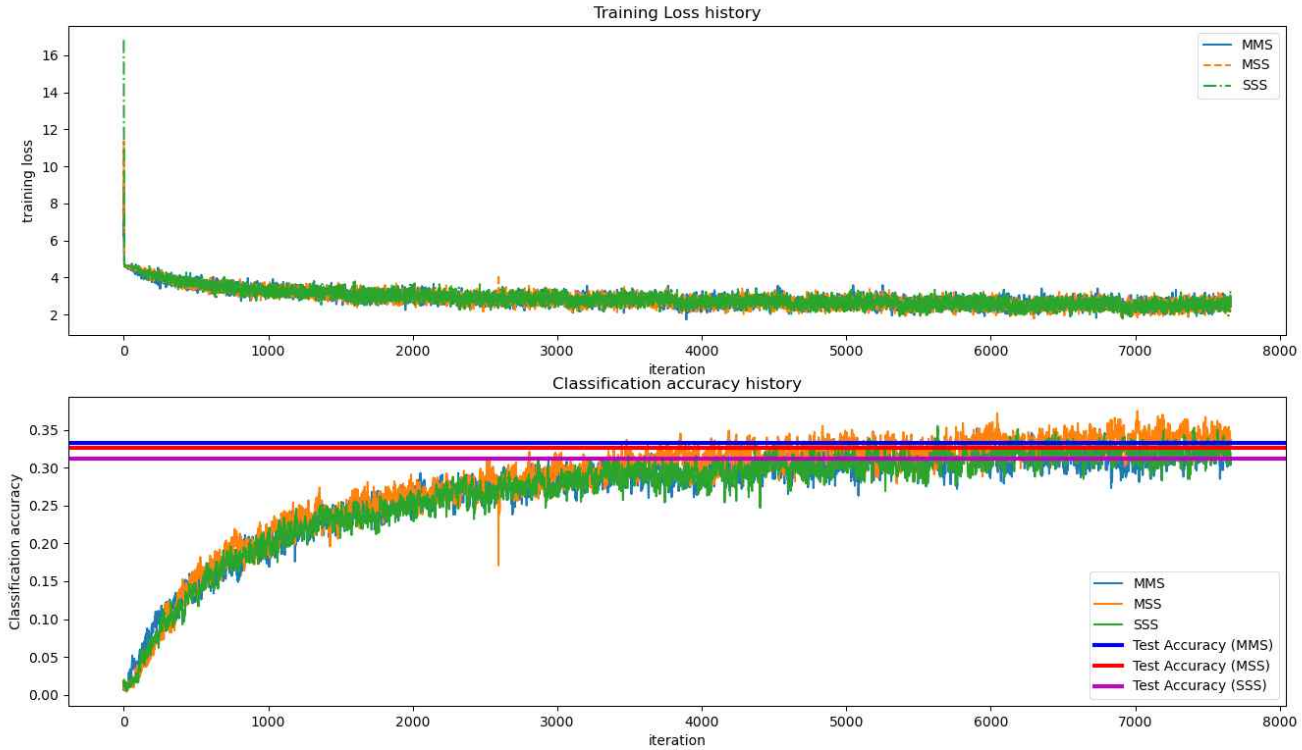

그림 6. [실험2] CIFAR-100 에 대한 Hybrid Pooling 결과: CIFAR-100 학습 데이터에 대해서, 두 풀링의 조합에 따른 손실 및 정확도 그래프. 풀링 A/B/C 순으로 각각 Max Pooling이면 'M'으로, Spectral Pooling이면 'S'로 표기하였다. 해당 그래프는 MMS, MSS, SSS 조합에 따른 손실 및 정확도 그래프이다.

4.2.1. 실험 2: 하이브리드(Hybrid) 풀링 조합에 따른 비교

해당 실험에서는 CIFAR-100에 대해서, 두 풀링 방법을 조합하여 실험해보았다. **그림4.**의 모델 구조에서, 풀링 A/B/C 순으로 각각 Max Pooling이면 'M'으로, Spectral Pooling이면 'S'로 표기하였다.

표 3. [실험2] Cifar-100에 대한 Hybrid Pooling 결과

Model Name	Test Accuracy	Remarks
MMS	0.3329	Max-Max-Spectral
MSS	0.3263	Max-Spectral-Spectral
SSS	0.3124	All Spectral Poolings
MMM	0.2806	All Max Poolings

표 3과 그림 6을 보면 모델의 앞단에 맥스 풀링을 달고, 뒷단에 Spectral Pooling을 달아놓은 하이브리드(Hybrid) 모델의 성능이 가장 좋은 것을 확인할 수 있다. 이는 **2.2.**에서 말한 바와 같이 층이 깊어질수록 세밀한 특징들을 잃는다는 맥스 풀링의 고질적인 문제점을 Spectral Pooling이 해결해주는 것으로 해석할 수 있다. 앞에 두 단까지는 맥스 풀링으로 특성맵에서 강한 부분들을 잘 뽑다가, 세 번째 단이 될 때 정보량을 많이 보존해주는 Spectral Pooling을 통해 최적화된 경로로 모델이 학습된 것이다.

5. 결론

본 논문에서는 DFT를 이용한 Spectral Pooling에 대한 소개와, 맥스 풀링과의 성능 비교 및 분석을 수행했다. 우선 학습 초반에 맥스 풀링보다 Spectral Pooling의 학습 속도가 빠르다는 결과를 얻을 수 있었다. 그리고 신경망이 깊어지면서 맥스 풀링이 정보 손실로 성능 저하를 초래할 때, 이런 상황을 방지하기 위해 Spectral Pooling을 상호보완적으로 사용할 수 있다는 가능성을 확인했다.

감사의 글

이 논문은 삼성전자 및 2020년도 BK21 플러스 창의정보기술 인재양성사업단에 의하여 지원되었음.

Reference

[1] A Krizhevsky, I Sutskever, GE Hinton. "Imagenet classification with deep convolutional neural networks." In Advances in Neural Information Processing Systems(NIPS), 2012.

[2] Y LeCun et al. "Handwritten digit recognition with a back-propagation network." In Advances in Neural Information Processing Systems(NIPS), 1990.

[3] N Akhtar, U Ragavendran. "Interpretation of intelligence in CNN-pooling process: A methodological survey." In Neural Computing and Applications, 2019.

[4] JT Springenberg et al. "Striving for simplicity: The all convolutional net." In eprint arXiv, 2014.

[5] O Rippel, J Snoek, RP Adams. "Spectral Representations for Convolutional Neural Networks." In Advances in Neural Information Processing Systems(NIPS), 2015.

[6] A Torralba, A Oliva. "Statistics of natural image categories." *Network*, 14(3):391-412, August 2003.