

정 20 면체 기반 360 도 이미지 표현 및 CNN 적용 방법

이연건^o, 정재석, 윤종섭, 조원준, 윤국진

한국과학기술원 기계공학과

{dldusrjs, jason.jeong, jseob, wonjune, kjyoon} @kaist.ac.kr

요 약

본 논문에서는 정 20 면체를 활용한 360 도 이미지 표현 방법과 정 20 면체에 대한 CNN 적용 방법을 제시한다. 정다면체 중 가장 높은 차원의 정 20 면체에 360 도 이미지를 투영하여 이미지의 왜곡을 최소화하고 정 20 면체의 삼각형 픽셀에 적용 가능한 convolution, pooling 방법을 고안함으로써 공간적 정보 손실없이, 기존 직사각형 이미지 기반 CNN 기법들의 360 도 이미지 활용 가능성을 검증한다. 먼저, 정삼각형을 4 개의 하위 정삼각형으로 분할하는 방법을 통해 정 20 면체를 적정 규모의 삼각형 픽셀로 표현하고 360 도 이미지 값을 투영한다. 그리고 각 삼각형 픽셀마다 인접한 픽셀과의 연결 관계를 정의하고 커널을 적용한다. 이를 통해, CNN 은 왜곡이 최소화된 360 도 이미지를 입력 받을 수 있으며, 공간적으로 균일한 정보를 추출할 수 있어 360 도 이미지에 대해 보다 높은 성능을 보인다.

1. 서론

CNN 은 이미지 분석에 효과적인 딥 러닝 기법으로, 최근 360 도 이미지에 적용하는 연구가 진행되고 있다. 대표적인 연구로 ERP[2], Cube map [1] 형태로 이미지를 변환하고 기존 CNN 기법을 적용하는 방식이 존재한다. 하지만 변환 과정 상 발생하는 이미지 왜곡과 공간적으로 커널이 균일하지 않게 적용되는 문제점이 있어 낮은 성능을 보인다. 본 연구에서는 360 도 이미지의 왜곡을 정 20 면체에 투영하는 방식으로 최소화하고 정 20 면체 표면에 균일하게 커널을 적용하여 360 도 이미지에 대한 CNN 성능을 향상시키는 방법을 제안한다. ERP, Cube map 방식들과 360 도 이미지 형태로 변환된 MNIST 와 SYNTHIA 데이터 셋에 대해 각각 classification 과 object detection 성능 실험을 진행하여 잠재성을 검증했다.

2. 이미지 표현 및 CNN 적용 방법

2.1 정 20 면체 기반 360 도 이미지 표현

360 도 이미지는 일반적으로 구 형태를 갖고 있다. 이미지를 CNN 의 입력 값으로 사용하기 위해서는 픽셀로 표현이 가능해야 하는데 모든 픽셀의 넓이가 일정하도록 구 표면을 픽셀 단위로 나누는 방법은 존재하지 않는다. 따라서 모든 픽셀의 넓이를 일정하게 유지하기 위하여 모든 면의 모양이 일정한 정다면체를 이용하되 가장 높은 차원이며, 구 형상에 가장 근접한 정 20 면체로 360 도 이미지를 표

현했다. 정 20 면체의 각 면은 정삼각형 형태로 1/4 크기의 정삼각형 4 개로 분할될 수 있는 특징을 갖고 있다. 20 개의 삼각형 면을 분할하여 매우 작은 크기의 삼각형으로 표현할 수 있고 각 삼각형은 pixel 로 간주된다. 360 도 이미지를 각 삼각형 픽셀의 face normal vector 를 기준으로 투영시켜 정 20 면체 이미지를 생성할 수 있다.

2.2 Convolution 기법

제안된 이미지 표현 방식에서 모든 삼각형 픽셀은 indexing 이 가능하며 각 픽셀마다 인접한 픽셀과의 연결 관계가 정의되어 있다. 이 연결 관계를 이용하여 픽셀마다 정해진 규칙에 따라 인접 픽셀을 호출함으로써 convolution 커널을 공간 상 균일하게 적용할 수 있다. 호출하는 인접 픽셀의 개수를 조절함으로써 다양한 크기의 커널을 적용할 수 있으며 커널의 크기는 반지름으로 설정할 수 있다. 그림 1 은 반지름 크기가 1 인 커널을 적용한 모습이다.

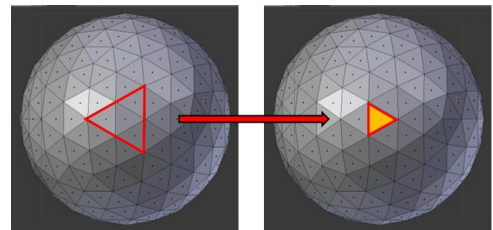


그림 1. Convolution 기법

2.3 Pooling 기법

Convolution 기법과 마찬가지로 각 픽셀 간의 연결 관계를 이용하여 인접 픽셀을 호출하고 max, average pooling 을 적용할 수 있다. 2.1 에서 언급한 삼각형 면 분할 방식의 역과정을 통해 4 개의 작은 삼각형을 1 개의 정삼각형으로 통합하는 과정을 반복하여 기존 CNN 의 pooling 과 동일한 의미를 갖도록 했다. 그림 2 는 1 번 분할된 상태에서 pooling 하는 과정이다.

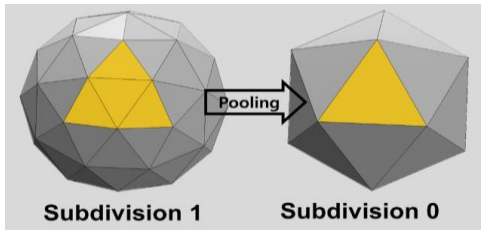


그림 2. Pooling 기법

3. 실험 결과 및 분석

제안된 방법의 성능을 평가하기 위하여, MNIST 와 SYNTHITA 데이터 셋을 360 도 이미지 형태로 변환하여 각각 classification 과 object detection 성능을 관찰했다. MNIST 이미지는 360 도 중 무작위로 선정한 위도와 경도에 투영하는 방식으로 변환했고 기존 ERP 이미지, Cube map 이미지 기반 기법들과 비교했다. SYNTHIA 이미지는 앞, 뒤, 좌, 우의 이미지를 연결하여 투영하는 방식으로 변환하고 mAP(%)를 계산했다.

표 1. MNIST classification 성능

	Correct Predictions	Test Set Size	Average accuracy(%)
ERP	528,577	700,000	75.51
Cube map	521,937	700,000	74.56
Ours	616,920	700,000	88.13

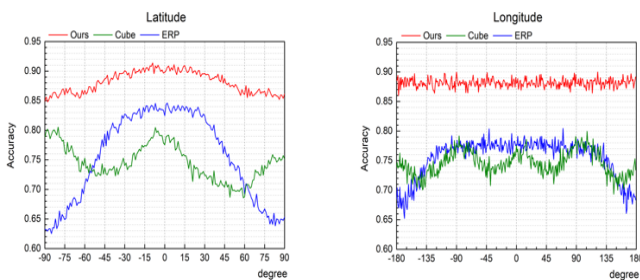


그림 1. 위도, 경도 별 MNIST classification 성능

먼저, 그림 1 은 classification 에 대해 공간상 이미지 왜곡이 주는 성능 차이를 관찰하기 위해 위도와 경도를 기준으로 성능을 정리한 결과이다. 기존 방식들이 이미지 왜곡이 심한 위도와 경도에서 성능 하락을 겪는 반면 제안된 기법은 가장 높

은 성능을 보일 뿐만 아니라 위도와 경도에 따라 균일한 성능을 보인다.

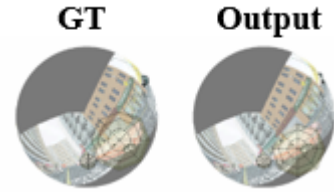


그림 4. SYNTHIA object detection 결과

표 2. SYNTHIA object detection 성능

	mAP(%)
Ours	17.3

그림 4 와 표 2 는 SYNTHIA object detection 을 수행한 결과이다. Ground truth 대비 객체의 크기와 위치를 검출해내는 것을 볼 수 있다. 또한, 지평선이 회전된 이미지에 대해서도 객체 검출을 수행할 수 있음을 관찰할 수 있다.

4. 결론

본 논문에서는 정 20 면체를 기반으로 한 360 도 이미지 표현 기법과 이를 CNN 에 적용하는 방법을 제안했다. 이미지 왜곡을 최소화함으로써 CNN 의 성능을 향상시킬 수 있음을 보였고 object detection 과 같은 난이도 높은 과제를 수행할 수 있음을 물론 360 도 이미지 내 지평선 회전에 강건함을 보이는 장점을 볼 수 있었다. 제안된 방법은 파라미터 규모 증가와 최적화 과정을 통해서 다양한 과제에 대한 360 도 이미지 처리 성능을 높일 수 있을 것으로 기대 된다.

감사의 글

이 연구는 삼성전자 미래기술육성센터의 지원을 받아 수행되었음. (과제번호 SRFC-TC1603-05).

참고문헌

[1] H.-T. Cheng, C.-H. Chao, J.-D. Dong, H.-K. Wen, T.-L. Liu, and M. Sun. Cube padding for weakly-supervised saliency prediction in 360 videos. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

[2] T. S. Cohen, M. Geiger, J. Kohler, and M. Welling. Spherical cnns. In arXiv preprint arXiv:1801.10130, 2018.