

이미지 매칭을 위한 디스크립터로서의 뉴럴 네트워크의 Feature Vector의 성능 분석

김도진
서강대학교 컴퓨터공학과
djkim3@sogang.ac.kr

Performance Analysis of Feature Vector of Neural Network as Descriptor for Image Matching

Dojin Kim

Department of Computer Science and Engineering

Sogang University

요 약

최근 급격하게 발전하고 있는 신경회로망 분야가 컴퓨터 비전 부분 연구에 많이 적용되고 있다. 이 신경회로망을 통해 계산된 이미지의 피쳐 벡터를 기존 사람이 구조를 직접 만든 구분자를 사용한 이미지 매칭 기술에 적용할 수 있을 것인지를 확인하기 위한 실험을 진행한다. 기존의 널리 사용되었던 SIFT 구분자와 VGGNet을 통해 추출된 피쳐 벡터를 사용해 이미지 매칭 실험을 진행하고, 그 결과에 대해 비교 분석해본다. 그리고 결과를 가지고 추후 진행할 연구에 대해서도 논의한다.

1. 서 론

이미지 매칭 기술은 컴퓨터 비전 분야의 기초적인 기술로 이미지 검색, 물체 인식, 얼굴 인식, 파노라마 스티칭 등의 분야에서 사용된다. 이미지 매칭 기술은 특징점 검출(Feature Detection), 특징점 구분자(Feature Descriptor) 생성, 특징점 매칭(Feature Matching)의 세 단계로 구분할 수 있다. 이미지 매칭의 기술로는 대표적으로 SIFT[1], SURF[2], ORB[3], BRISK[4]가 존재한다. 최근 이미지 데이터의 양이 급격하게 증가하고, 그래픽 처리 장치의 발전으로 신경 회로망에 대한 연구가 활발하게 진행되고 있다. 이를 통해 이미지 처리의 많은 분야, 예를 들어 이미지 검출 같은 영역에서는 고전적인 방법을 사용했을 때의 성능을 앞질러 사람의 능력을 뛰어넘는 성능을 보여주고 있다.

본 논문에서는 이미지 매칭에서 현재 활발하게 사용되고 있는 SIFT나 BRISK같은 사람이 구조를 직접 만든 특징점 구분자와 신경회로망을 통해 구성된 피쳐 벡터를 사용해 이미지 매칭의 성능을 분석하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 신경회로망과 관련한 연구들을 살펴본다. 3장에서는 이미지 매칭을 위한 구분자로서 신경회로망의 피쳐 벡터를 어떻게 사용할 것인지를 살펴보고, 피쳐 벡터를 사용한 이미지 매칭의 실험과 결과를 진행한다. 4장에서는 결론과 추후 개선될 사안에 대해 논의한다.

2. 관련 연구

2.1 신경 회로망

신경회로망(Neural Network)은 생물학적 뇌의 작동원리를 그대로 모방하여 문제를 해결하는 방법이다. 간단하

게 입력과 출력 사이에 은닉 노드(hidden node)들을 만들어 입력과 은닉 노드, 은닉 노드와 출력 사이의 가중치를 학습하는 방법이다. 그림 1은 기본적인 3-레이어 신경망을 나타낸 것이다. 회로망 내의 노드들은 어떻게 구성하는가에 따라 다양한 신경 회로망이 만들어진다. 대표적인 신경 회로망으로는 AlexNet[5], VGGNet[6], ResNet[7], googLeNet[8] 등이 있다.

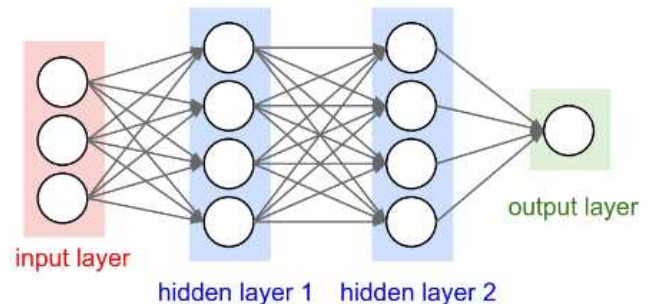


그림 1 . 기본적인 3-레이어 신경망

3. VGGNet과 신경회로망의 피쳐 벡터

3.1 VGGNet

VGGNet은 K. Simonyan과 A. Zisserman가 2014년 ILSVRC에서 발표한 합성곱 신경회로망(Convolutional Neural Network)이다. 합성곱 신경회로망은 여러 개의 비선형적인 특성을 가진 합성곱 층과 층의 크기를 변화시키는 샘플링 층으로 구성된 신경회로망이다. 그림 2는 VGGNet의 구조를 나타낸다. 이 신경회로망에서 모든 합성곱은 3x3, 풀링은 2x2만으로 이뤄져있다. 이렇게 구조가 간단하기 때문에 신경회로망을 쉽게 변형을 시켜가면서 여러 가지 구조에 대한 테스트를 할 수 있는 장점이 있다.



그림 2. VGGNet의 레이어 구조

3.2 신경회로망의 피쳐 벡터

이미지 분류를 위한 신경회로망은 회로망의 최종 단계에서 클래스에 대한 점수를 출력한다. 이 클래스에 대한 점수를 계산하는데 입력으로 들어오는 레이어를 입력 이미지의 정보를 포함하고 있는 피쳐 벡터라고 볼 수 있다. 이 피쳐 벡터는 이미지에 대한 정보를 포함하고 있기 때문에 구분자로서 사용이 가능 할 것이다. 그래서 이 피쳐 벡터가 기존의 사람이 직접 구조를 정해서 계산한 구분자와 비교했을 때, 어떤 매칭 성능을 낼 수 있을 지 실험해보기로 한다.

4. 실험

4.1 VGG-m Net

이미지 매칭에 사용하는 신경회로망은 2013년 ILSVRC의 우승한 Zeiler와 Fergus의 신경망을 VGGNet의 저자가 업데이트한 VGG-m Net[9]이다. 그림 3은 VGG-m Net의 구조이다. 더 작은 Receptive Window 크기를 갖고 있는 것이 특징이다. 그림 4와 같이 fc7 레이어의 크기를 원래의 4096에서 128, 1024, 2048로 변화를 주어 피쳐 벡터의 크기가 변화됨에 따라 매칭 성능에 어떤 영향을 주는 지 확인하였다. 신경회로망은 이미지 넷 2014 데이터로 학습되었다.

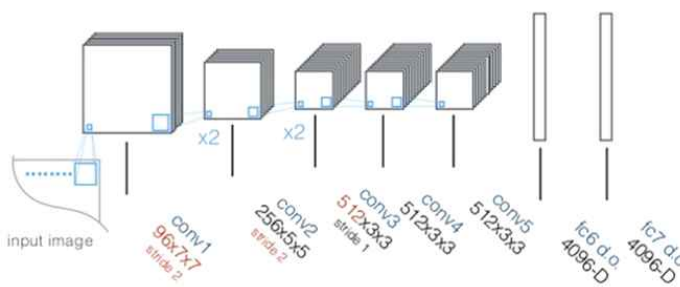


그림 3. VGG-m Net의 구조

4.2 실험 환경

이미지 매칭에 사용된 데이터는 특징점 추출, 특징점 구분자 및 이미지 매칭 기법에 대한 성능 평가에 대한 많은 실험에서 사용되는 Mikolajczyk의 데이터베이스를 사용하였다. Mikolajczyk 데이터 베이스는 여러 이미지의 변화 특성을 적용한 각각 6개의 이미지를 가지고 있는 8개의 세부 데이터 베이스 세트들로 구성되어 있다. 이미지에 적용된 변화는 흐림(Blur) 변화, 관찰 시점(Viewpoint) 변화, 크기 및 회전(Zoom, Rotation) 변화, 밝

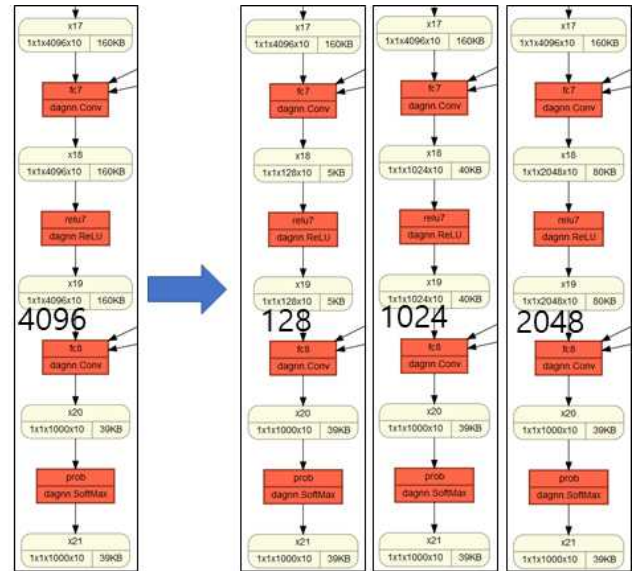


그림 4. VGG-m Net에서 크기를 변화시킨 레이어

기(Light) 변화, JPEG 압축(JPEG Compression) 변화 등이다.

Mikolajczyk 데이터 베이스에서 가우시안 차(Difference of Gaussian, DoG) 특징점 검출방법으로 검출한 특징점의 위치에서 96x96 크기의 이미지 패치를 추출하여 입력 이미지로 사용하였다. 1개의 이미지 당 1000개의 이미지 패치를 추출하였다.

매칭 판단 조건은 1번 이미지의 이미지 패치1의 피쳐 벡터와 2번 이미지의 모든 이미지 패치의 피쳐 벡터간의 L2 거리(L2 Distance)를 계산하고, L2 거리 중 가장 가까운 거리와 그 다음 가까운 거리의 비율이 0.8 미만이면 두 이미지 패치가 매칭이라고 한다.(SIFT 구분자 이미지 매칭 실험에서의 매칭 조건이다.)

다른 신경회로망과의 결과를 비교해 보기 위해서 ResNet과 googLeNet의 피쳐 벡터로 사용해 실험 하였다.

4.3 실험 결과 및 분석

표1은 이미지 매칭 실험 결과이다. 각 이미지 세트에 대한 모델의 이미지 매칭 개수를 나타낸 것이다. 가장 높은 매칭 개수를 굵은 글씨로 표시하였다.

실험 전 SIFT 구분자의 차원의 크기인 128과 같은 크기로 변화시킨 VGG-m-128 Net의 피쳐벡터가 가장 좋은 성능을 보여줄 것이라고 예측한 것과는 반대로 결과를 보면 피쳐의 크기가 4096인 VGG-m Net이 SIFT보다 더 많이 매칭시킨 이미지 세트를 확인할 수 있다. 특정 변화를 가진 이미지 세트에서 VGG-m Net의 매칭개수가 높게 나온 것이 아니어서 어떤 변화에 대해 신경회로망의 피쳐 벡터가 매칭 성능이 높다고 특정할 수는 없지만, SIFT의 결과보다 높은 성능을 보였다.

이를 통해 이미지 매칭에 신경회로망의 피쳐 벡터가 조금의 수정을 거치면 기존의 사람이 직접 구조를 만들었던 구분자들을 대체해 유용하게 쓰일 수 있을 것이라는 것을 확인할 수 있었다.

표1. 구분자와 신경회로망을 통한 피쳐 벡터를 사용한 이미지 매칭 실험 결과

	SIFT	vgg-m	vgg-m -2048	vgg-m -1024	vgg-m -128	ResNet	googLeNet
BARK	161.4	181	181.2	166.2	150.6	129.8	155.6
BIKES	325	305.4	289.8	166.2	270.6	227.4	276.6
BOAT	226	249.6	229.4	225.6	210.2	175.4	199.6
BRICKS	227.8	121.6	118.8	113	97	119.4	104.6
CARS	424.2	308.2	293.2	286.6	270.8	235.2	271.2
GRAFFITI	173.6	253.4	249.4	238.2	223.6	166	221.8
TREES	107.4	177.4	172.8	166.8	142	149.8	149.6
UBC	457.2	451	450.4	445.6	424	414.4	388

5. 결론 및 추후 연구

이번 실험을 통해 신경회로망을 통해 추출된 이미지의 피쳐 벡터가 이미지 매칭 분야에서 사람이 직접 구조를 구성하였던 기존의 구분자들을 대체할 수 있을 것이라는 가능성을 확인할 수 있었다. 특정한 이미지의 변화에 대해 신경회로망의 피쳐 벡터의 매칭 결과가 우수하게 나타난 것이 아니라서 어떤 영향을 주는지 더 연구가 필요할 것이라고 보인다. 그리고 이번 연구는 기존의 대량의 데이터로 학습된 회로망에서 추출된 피쳐 벡터가 이미지 매칭에 사용될 수 있을 것인지를 확인한 것이어서 특징점 검출 방법을 통해 추출된 이미지 패치에 대해 학습하지 않았다. 추후에는 이미지 패치에 대해 학습된 신경회로망을 통한 피쳐 벡터가 이미지 매칭에 어떠한 성능을 보여주는지 실험을 진행할 것이다.

참 고 문 헌

- [1] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision (IJCV), vol. 60, no. 12, pp. 91-110, 2004.
- [2] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "SURF: Speeded Up Robust Features," Computer Vision and Image Understanding (CVIU), vol. 110, no. 3, pp. 346-359, 2008.
- [3] E. Rublee, V. Rabaud, K. Konolige and G. R. Bradski, "ORB: An Efficient Alternative to SIFT or SURF," in Proc. IEEE International Conference on Computer Vision (ICCV), pp. 2564-2571, 2011.
- [4] S. Leutenegger, M. Chli and R. Y. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints," in Proc. IEEE International Conference on Computer Vision (ICCV), pp. 2548-2555, 2011.
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Proc. Neural Information Processing Systems(NIPS), vol. 1, pp. 1097-1105, 2012.
- [6] K. Simonyan, and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014
- [7] K. He, X. Zhang, S. Ren, and J. Sun. "Delving deep into rectifiers: Surpassing human-level performance on imagenet in Proc. IEEE International Conference on Computer Vision (ICCV), 2015
- [8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proc. Computer Vision and Pattern Recognition(CVPR), 2015
- [9] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the Devil in the Details: Delving Deep into Convolutional Networks," British Machine Vision Conference(BMVC) ,2014