Convolutional Neural Network (CNN)

Prof. Ph.D. Woo Youn Kim

Chemistry, KAIST

Goals

|			
5주	주제	Multilayer perceptron 2	
	목표	Knowing various issues on MLP and techniques to resolve them	
	내용	Overfitting, regularization, dropout, batch normalization, cross validation	
6주	주제	Convolutional Neural Network (CNN) & SMILES	
	목표	Understanding CNN and molecular representation with SMILES	
	내용	Convolution, receptive field, stride, pooling Supervised learning of Log P and TPSA	
7주	주제	Molecular graphs & Graph Neural Network (GNN)	
	목표	Understanding GCN and molecular representation with graphs	
	내용	Molecular graph representation, graph convolutional network Supervised learning of logP and TPSA	
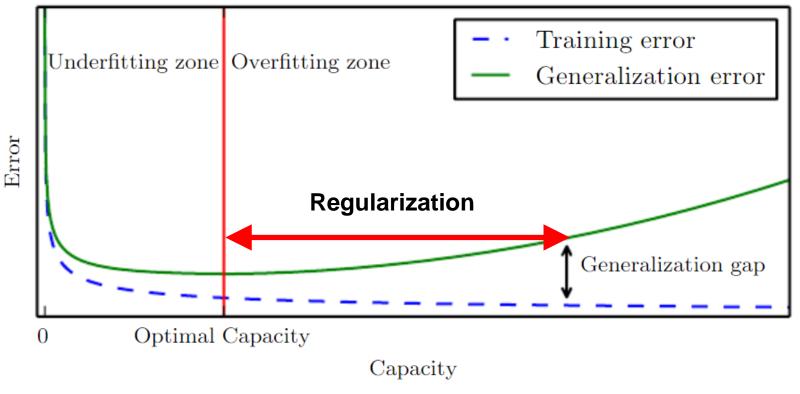
Contents

- Molecular representation with SMILES

- Convolutional neural network (CNN)

- Receptive field & Convolution

- Padding

- Skip-connection

- CNN in chemistry

- Practice session: supervised learning of log P and TPSA:

Review

- The main challenge is to find an **optimal capacity** of model for a given task.

- **Regularization** is any modification we make to a learning algorithm that is intended to reduce its generalization error but not its training error.

CNN, RNN, GCN

Source

Deep Learning

An MIT Press book

Ian Goodfellow and Yoshua Bengio and Aaron Courville

Exercises   Lectures   External Links

The Deep Learning textbook is a resource intended to help students and practitioners enter the field of machine learning in general and deep learning in particular. The online version of the book is now complete and will remain available online for free.

The deep learning textbook can now be ordered on Amazon.

https://www.deeplearningbook.org/

KAIST EE Jinwoo Shin
http://alinlab.kaist.ac.kr/ee807_2018.html

http://cs231n.github.io/convolutional-networks/

Reference paper

DeepDTA: deep drug–target binding affinity prediction

Hakime Öztürk[1], Arzucan Özgür[1,*] and Elif Ozkirimli[2,*]

[1]Department of Computer Engineering and [2]Department of Chemical Engineering, Bogazici University, Istanbul 34342, Turkey

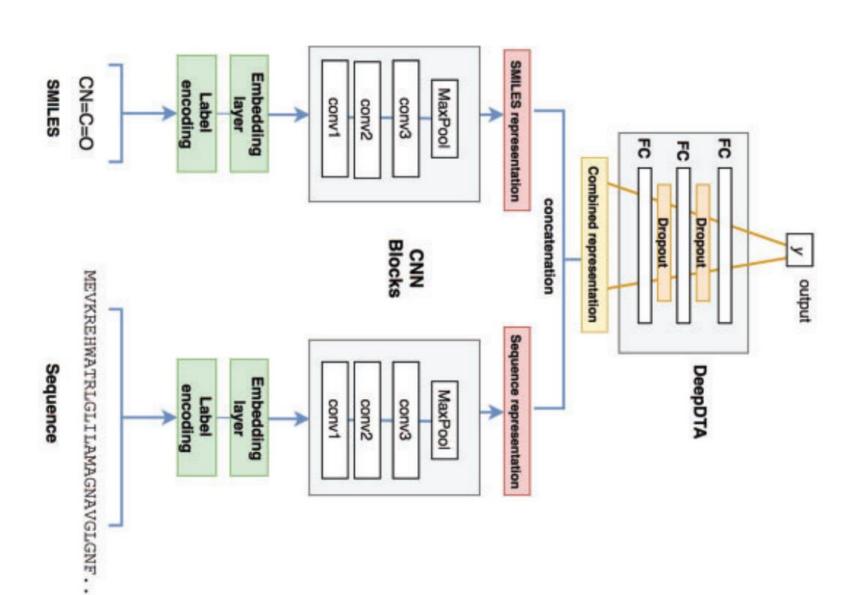*To whom correspondence should be addressed.

Reference paper

Abstract

Motivation: The identification of novel drug–target (DT) interactions is a substantial part of the drug discovery process. Most of the computational methods that have been proposed to predict DT interactions have focused on binary classification, where the goal is to determine whether a DT pair interacts or not. However, protein–ligand interactions assume a continuum of binding strength values, also called binding affinity and predicting this value still remains a challenge. The increase in the affinity data available in DT knowledge-bases allows the use of advanced learning techniques such as deep learning architectures in the prediction of binding affinities. In this study, we propose a deep-learning based model that uses only sequence information of both targets and drugs to predict DT interaction binding affinities. The few studies that focus on DT binding affinity prediction use either 3D structures of protein–ligand complexes or 2D features of compounds. One novel approach used in this work is the modeling of protein sequences and compound 1D representations with <u>convolutional neural networks (CNNs).</u>
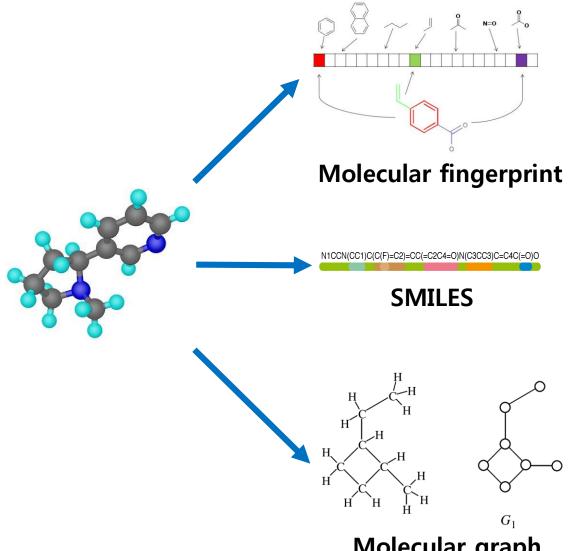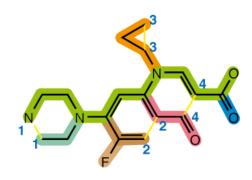
Reference paper

Molecular representation

Molecular representation in deep learning

Molecular fingerprint

N1CCN(CC1)C(C(F)=C2)=CC(=C2C4=O)N(C3CC3)C=C4C(=O)O
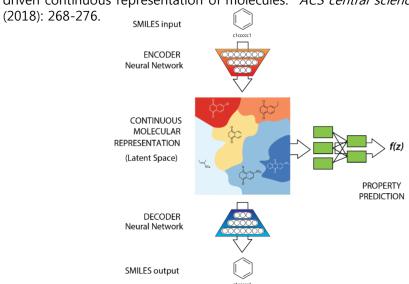
SMILES

Molecular graph

SMILES

SMILES : 1D string
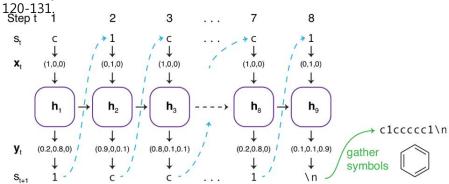

N1CCN(CC1)C(C(F)=C2)=CC(=C2C4=O)N(C3CC3)C=C4C(=O)O

✓ Simply match a molecule to string

✓ Easy to construct database

✓ Can use natural language AI models

Gómez-Bombarelli, Rafael, et al. "Automatic chemical design using a data-driven continuous representation of molecules." *ACS central science* 4.2 (2018): 268-276.

Segler, Marwin HS, et al. "Generating focused molecule libraries for drug discovery with recurrent neural networks." *ACS central science* 4.1 (2017): 120-131.
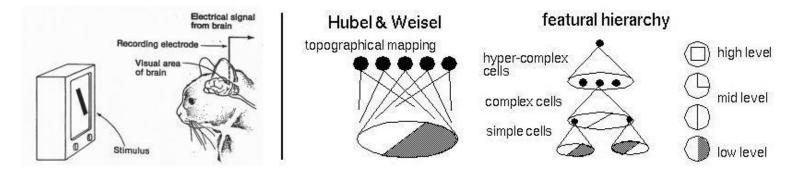
Introduction to CNN

History

Work by Hubel and Wiesel in the 1950s and 1960s showed that cat and monkey visual cortexes contain neurons that individually respond to small regions of the visual field.

The **receptive field** of an individual sensory neuron is the particular region of the sensory space (e.g., the body surface, or the visual field) in which a stimulus will modify the firing of that neuron.
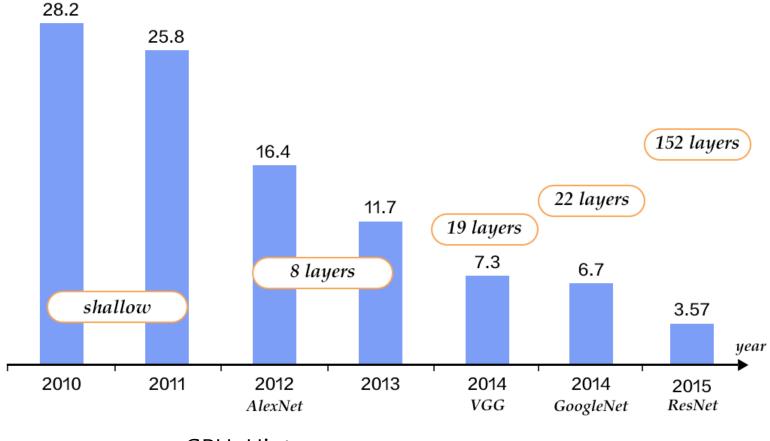
LeCun et al. in 1989, used back-propagation to learn the convolution kernel coefficients directly from images of hand-written numbers.

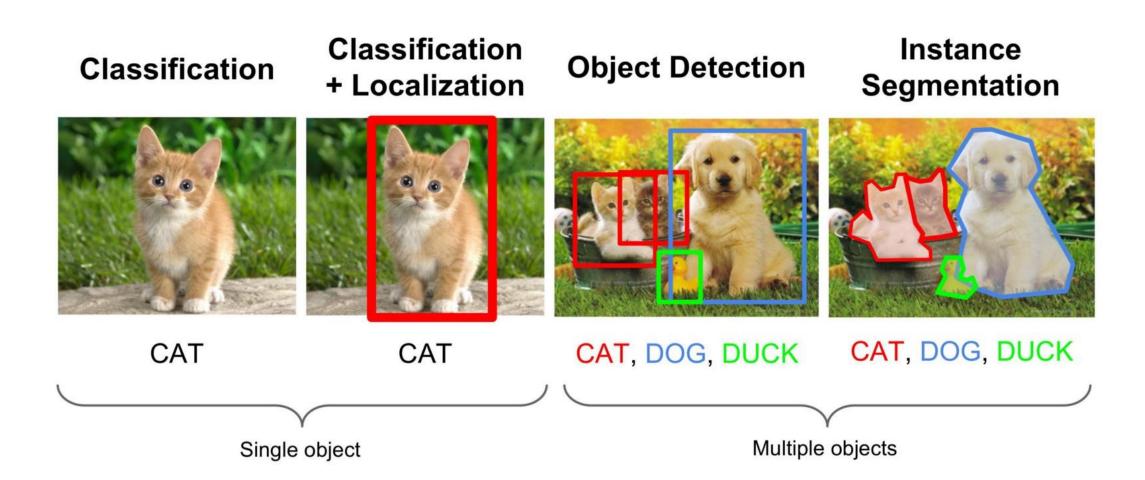LeNet-5, a pioneering 7-level convolutional network by LeCun et al. in 1998

ILSVRC Winners
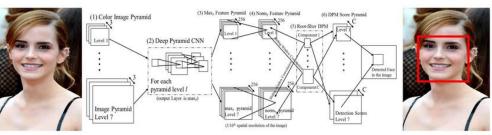
ImageNet Large Scale Visual Recognition Challenge

28.2 — 2010 — *shallow*
25.8 — 2011
16.4 — 2012 AlexNet — *8 layers*
11.7 — 2013
7.3 — 2014 VGG — *19 layers*
6.7 — 2014 GoogleNet — *22 layers*
3.57 — 2015 ResNet — *152 layers*

year

GPU, Hinton group

CNN

Classification	Classification + Localization	Object Detection	Instance Segmentation
CAT	CAT	CAT, DOG, DUCK	CAT, DOG, DUCK

Single object — Multiple objects

https://medium.com/comet-app/review-of-deep-learning-algorithms-for-object-detection-c1f3d437b852

Applications of CNN

Image recognition
       Facebook, Apple, MS, etc

Video analysis
       Commentary, Self-driving car, etc
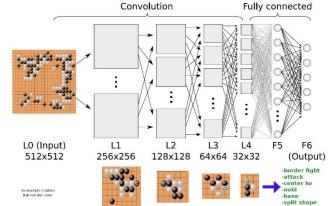
Natural language processing
       Translation, automatic captioning, etc

Drug discovery: AtomWise

Games: DeepMind, Go, Atari, etc

Applications of CNN

삼성전자 뛰쳐나와 만든 앱, 찍는 순간 칼로리가 나온다
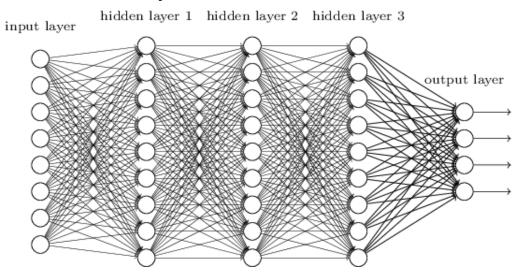
[출처: 중앙일보] 삼성전자 뛰쳐나와 만든 앱, 찍는 순간 칼로리가 나온다

https://news.joins.com/article/22973973

CNN

Drawbacks of DNN

- Previous DNNs (MLPs) use fully-connected layers
- Connect all the neurons between the layers

Drawbacks

Large number of parameters

- Easy to be over-fitted

Large memory consumption (GPU issues)

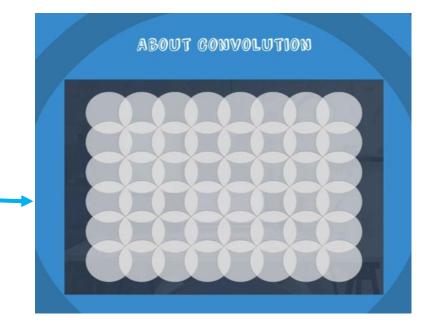Does not enforce any structure, e.g., local information

In many applications, local features are important, e.g., images, language, etc.

Key idea of CNN

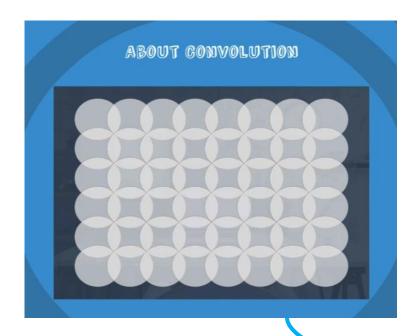**Weight sharing** with a **receptive field** and **convolving** it over inputs

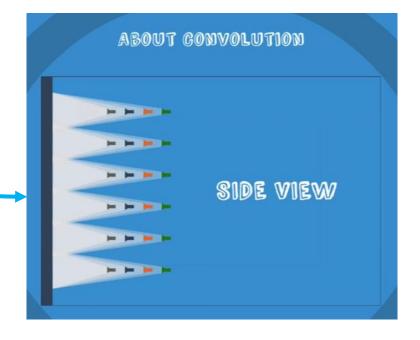- **Each receptive field (filter) shines on specific area of picture and convolve over inputs**

Key idea of CNN

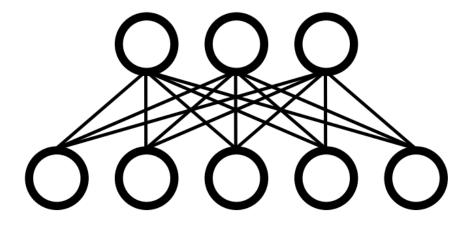**Weight sharing** with a **receptive field** and **convolving** it over inputs

- **Each receptive field (filter) shines on specific area of picture and convolve over inputs**
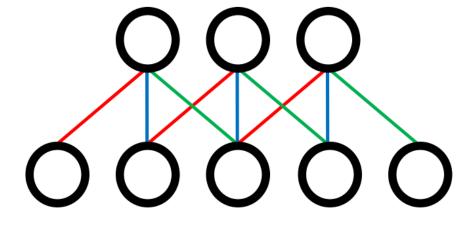- **Four different filters shine each area and try to detect patterns.**

https://www.youtube.com/watch?v=JiN9p5vWHDY&index=8&list=PLjJh1vlSEYgvGod9wWiydumYl8hOXixNu

Key idea of CNN

Weight sharing with a **receptive field** and **convolving** it over inputs

DNN (fully connected)   CNN (weight sharing and convolving)

- Reduce the number of parameters (less overfitting)
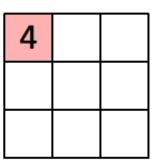
- Learn local features

- Translation invariance

Receptive field
& Convolution

Receptive field

Apply same weights over the different spatial regions

Image

Convolved
Feature

Local feature

$$h_k^{(l+1)} = \sigma(\sum_{ij \in Fk} w_{ij}^{(l)} h_{ij}^{(l)} + b^{(l)})$$
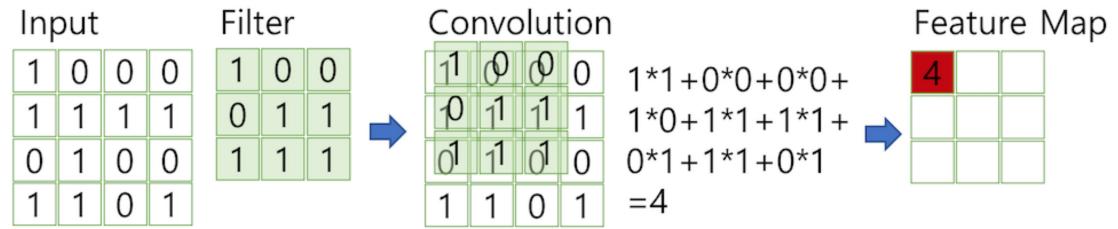
A selection of input maps by the receptive field

Receptive field

Input			
1	0	0	0
1	1	1	1
0	1	0	0
1	1	0	1

Filter		
1	0	0
0	1	1
1	1	1

Convolution

1	0	0	0
0	1	1	1
0	1	0	0
1	1	0	1

1*1+0*0+0*0+
1*0+1*1+1*1+
0*1+1*1+0*1
=4

Feature Map

4		

http://taewan.kim

$$h_k^{(l+1)} = \sigma(\sum_{ij \in Fk} w_{ij}^{(l)} h_{ij}^{(l)} + b^{(l)})$$

Example

Input: 5×5
Receptive field (or filter): 3×3
Stride: 1
Pooling: summation
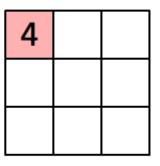
Image
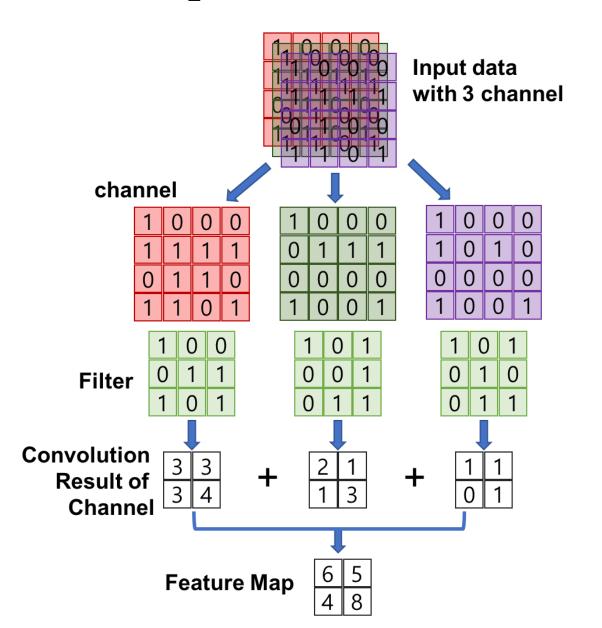
Convolved Feature
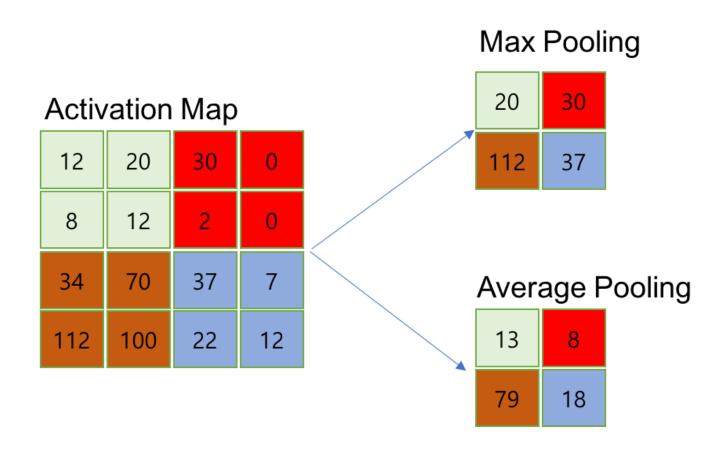
Multiple channels

Pooling

Activation Map

12	20	30	0
8	12	2	0
34	70	37	7
112	100	22	12

Max Pooling

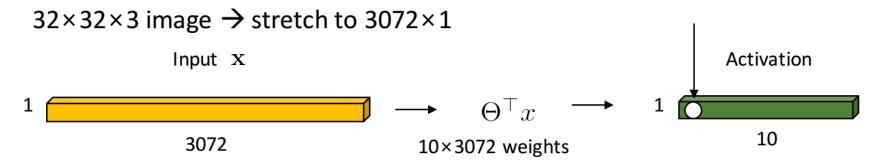| 20	30
112	37

Average Pooling

13	8
79	18

- Make the representations smaller

- Reduce number of parameters and computation ➔ Regularization effect
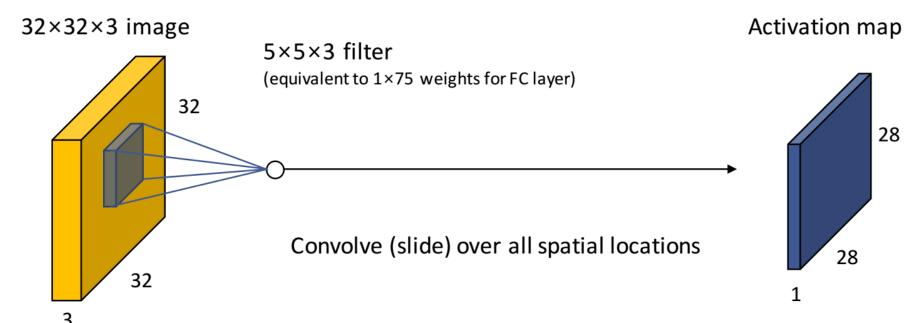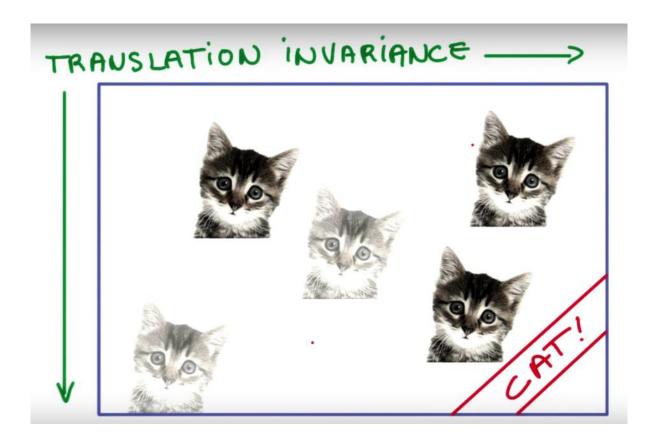
Receptive field

- ## Fully connected layer

$32 \times 32 \times 3$ image $\rightarrow$ stretch to 3072×1

Input x

Activation

1

3072

$\Theta^\top x$

10×3072 weights

1

10

- ## Convolution layer

$32 \times 32 \times 3$ image

$5 \times 5 \times 3$ filter
(equivalent to 1×75 weights for FC layer)

32

32

3

Activation map

28

28

1

Convolve (slide) over all spatial locations

Translation invariance

When input is translated, the corresponding output to recognize the object should not be changed.

CNN can produce the same output even though the input image is shifted due to weight sharing.

Padding

Padding

9

0	0	0	0	0	0	0	0	0
0								0
0								0
0								0
0								0
0								0
0								0
0								0
0	0	0	0	0	0	0	0	0

9

7×7 input (spatially)
Zero pad 1 pixel border
Assume 3×3 filter
Applied with **stride 3**

→ **3×3 output**

Formula for output size

- Input data hieght: H
- Input data width: W
- Filter height: FH
- Filter width: FW
- Stride size: S
- Padding size: P

$$OutputHeight = OH = \frac{(H + 2P - FH)}{S} + 1$$

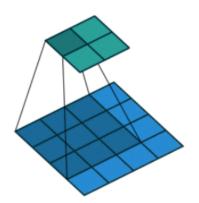$$OutputWeight = OW = \frac{(W + 2P - FW)}{S} + 1$$

Formula for output size

- Output size must be integer

- If a pooling layer coming after conv layer, the size of feature map should be integer times of the pooling size

   Ex) Pooling size: 3x3
      Then, the output size before the pooling may be 9x9 or 9x6.

- FH, FW, S, P, and Pooling size should be determined in such a way

- Input data hieght: H

- Input data width: W

- Filter height: FH

- Filter width: FW

- Stride size: S

- Padding size: P

$$Output Height = OH = \frac{(H + 2P - FH)}{S} + 1$$
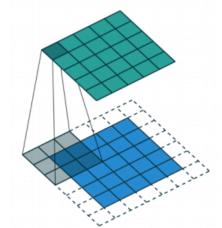
$$Output Weight = OW = \frac{(W + 2P - FW)}{S} + 1$$
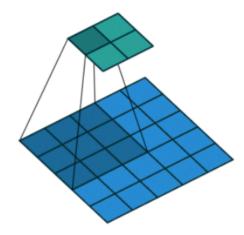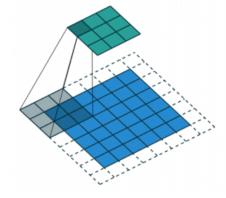
More examples
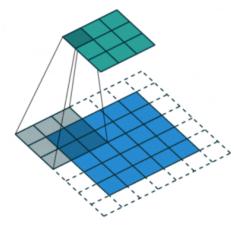
No padding, stride 1

Padding 1, stride 1
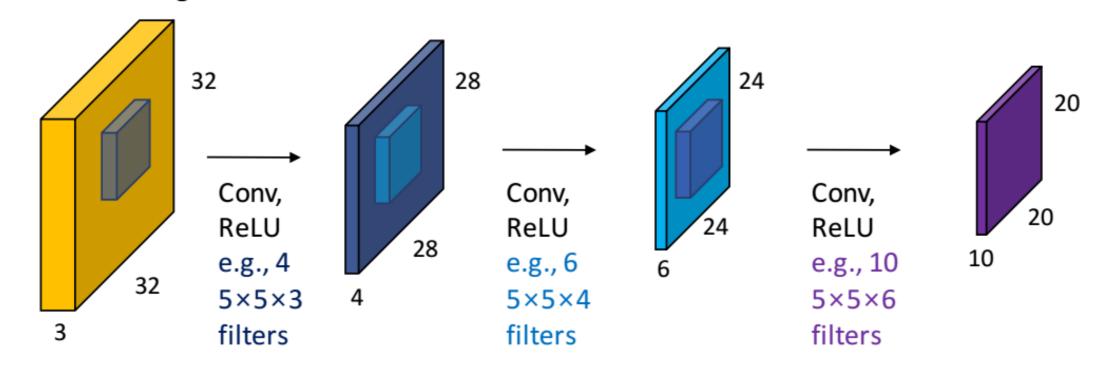
Padding 1, stride 2 (odd)

No padding, stride 2

Padding 1, stride 2

ConvNet

32×32×3 image

32

32

3

→

Conv,
ReLU
e.g., 4
5×5×3
filters

28

28

4

→

Conv,
ReLU
e.g., 6
5×5×4
filters

24

24
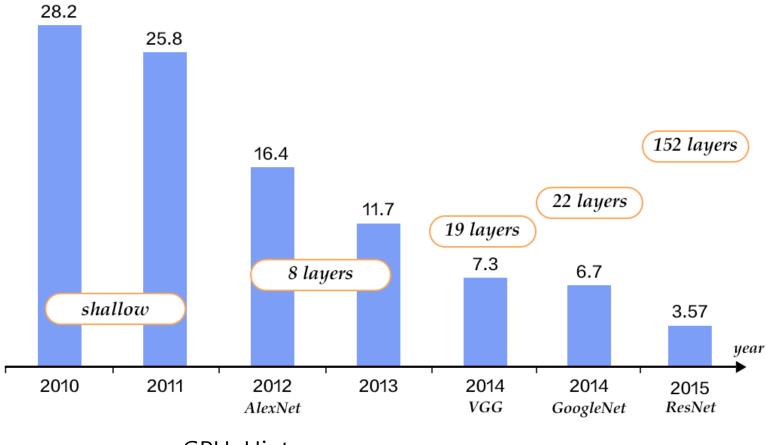
6

→

Conv,
ReLU
e.g., 10
5×5×6
filters

20

20

10

Technical issues

Going wider: inception
Going deeper: skip connection

ILSVRC Winners

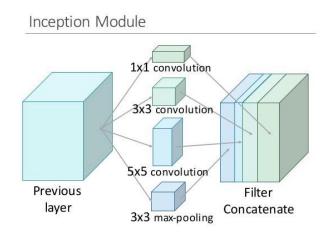**ImageNet Large Scale Visual Recognition Challenge**

GPU, Hinton group

Inception

- To go **wiiiiiider**

- To extract different features
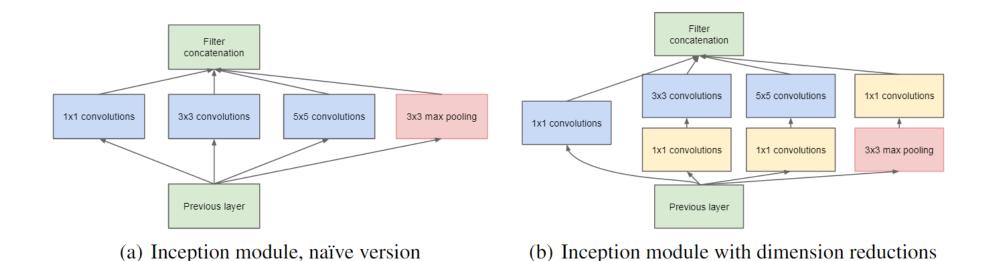

Each layer type extracts a different feature.
The output of a 5x5 receptive field is different from the output of a 3x3 one.

Computes *multiple different transformations* over the same input map in parallel.

Then, the network chooses the best combination.

https://towardsdatascience.com/an-intuitive-guide-to-deep-network-architectures-65fdc477db41

Inception

(a) Inception module, naïve version
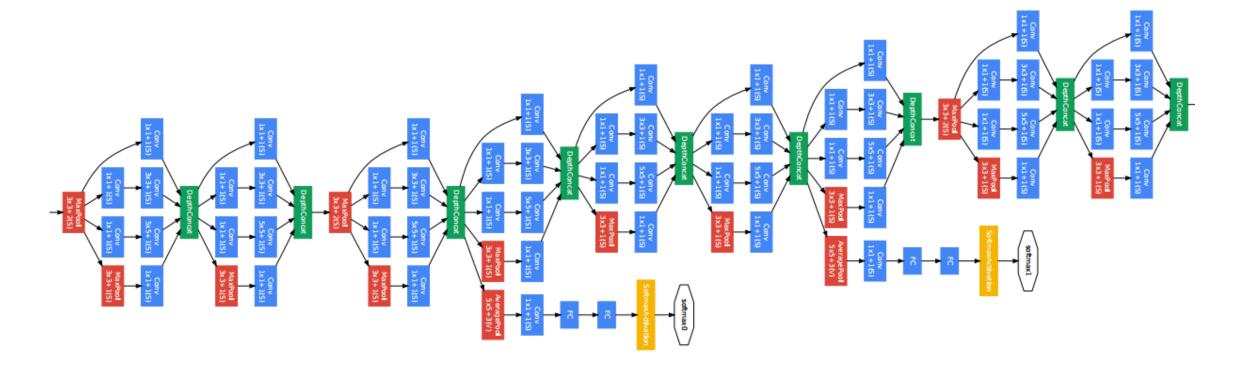
(b) Inception module with dimension reductions

Stack different layer transformations in parallel, resulting in nets that were **simultaneously deep** (many layers) and **wide** (many parallel operations).
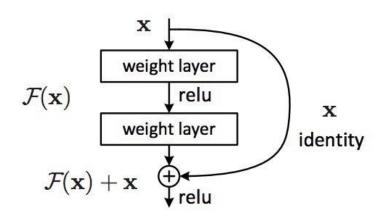
https://towardsdatascience.com/an-intuitive-guide-to-deep-network-architectures-65fdc477db41

Inception

GoogLeNet – Winner of 2014 ImageNet Challenge

Skip connection

- To go **deeeeeeper**

- To avoid the **vanishing gradient**

Each "block" in ResNet consists of a series of layers and a skip connection adding the input of the block to its output.
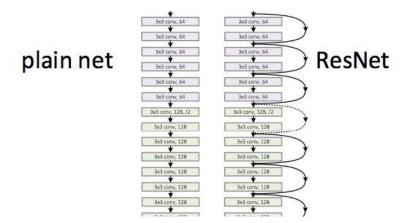
The "add" operation is performed element-wise
If the input and output are of different sizes, zero-padding or projections (via 1x1 convolutions) can be used to create matching dimensions.

Skip connection

ResNet (Residual net) – Winner of 2015 ImageNet Challenge

The gradient signal in ResNets could travel back directly to early layers via skip connections
➔ Going deeper and deeper to 50-layer, 101-layer, 152-layer, and even 1000+ layer nets.
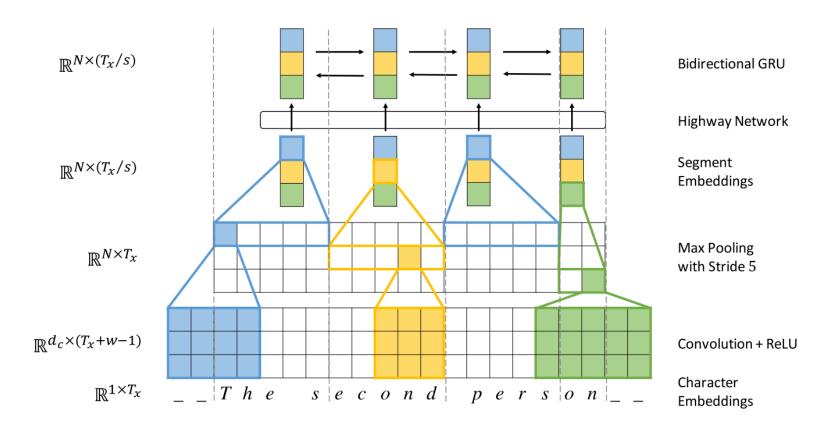
Examples

NLP example

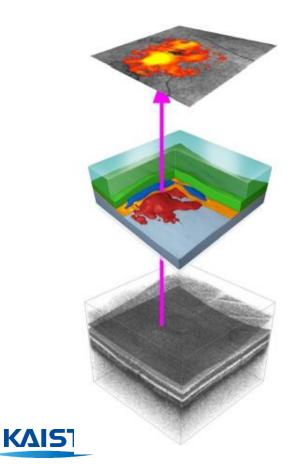Figure 1: Encoder architecture schematics. Underscore denotes padding. A dotted vertical line delimits each segment.
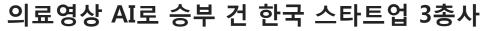
Lee, Jason, Kyunghyun Cho, and Thomas Hofmann. "Fully character-level neural machine translation without explicit segmentation." *arXiv preprint arXiv:1610.03017* (2016).

Medical imaging

DeepMind

Clinically applicable deep learning for diagnosis and referral in retinal disease

의료영상 AI로 승부 건 한국 스타트업 3총사

Lunit cancer

VUNO Bone age

DEEP BIO cancer
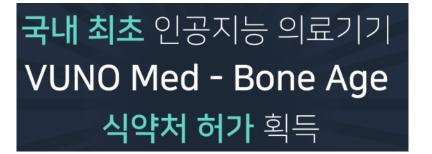
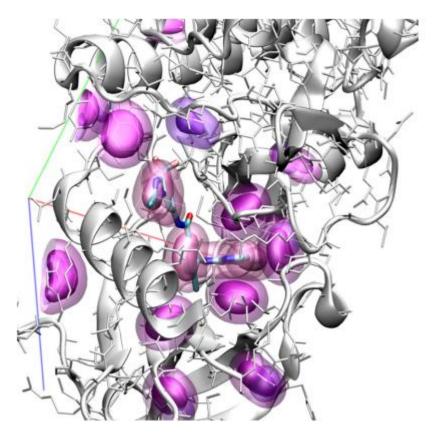국내 최초 인공지능 의료기기
VUNO Med - Bone Age
식약처 허가 획득

CNN in Chemistry

Protein−Ligand Scoring with Convolutional Neural Networks

Matthew Ragoza,[†,‡] Joshua Hochuli,[‡,¶] Elisa Idrobo,[§] Jocelyn Sunseri,[∥] and David Ryan Koes*[,∥]

[†]Department of Neuroscience, [‡]Department of Computer Science, [¶]Department of Biological Sciences, and [∥]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States
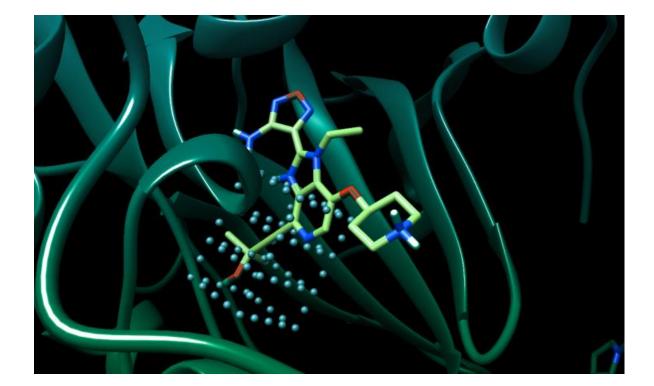[§]Department of Computer Science, The College of New Jersey, Ewing, New Jersey 08628, United States

We describe convolutional neural network (CNN) scoring functions that take as input a comprehensive three-dimensional (3D) representation of a protein−ligand interaction.

Visualization of atom densities used as input to CNN scoring. Aromatic carbon atom densities are shown at two isosurface levels (solid and transparent surfaces) for both the receptor (purple) and ligand (lavender).

our CNN scoring function outperforms the AutoDock Vina scoring function when ranking poses both for pose prediction and virtual screening.

AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery

Atomwise
Better medicines faster.

Izhar Wallach
Atomwise, Inc.
izhar@atomwise.com

Michael Dzamba
Atomwise, Inc.
misko@atomwise.com

Abraham Heifets
Atomwise, Inc.
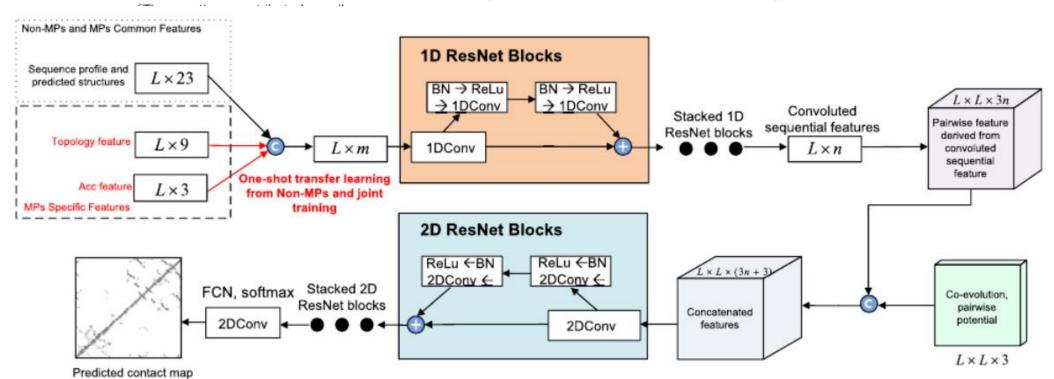abe@atomwise.com

more than $51 million in funding

Folding Membrane Proteins by Deep Transfer Learning

Computational elucidation of membrane protein (MP) structures

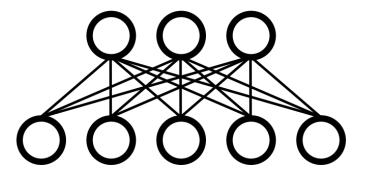Sheng Wang,[1,2,4,5] Zhen Li,[1,3,5] Yizhou Yu,[3] and Jinbo Xu[1,6,*]

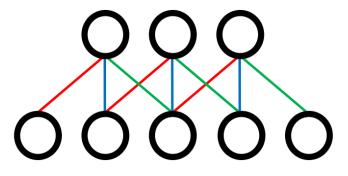[1]Toyota Technological Institute at Chicago, Chicago, IL 60637, USA
[2]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
[3]Department of Computer Science, University of Hong Kong, Hong Kong
[4]Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Summary

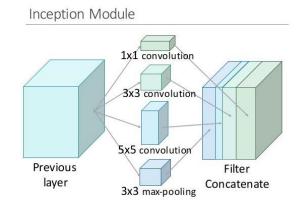**Weight sharing** with a **receptive field** and **convolving** it over inputs
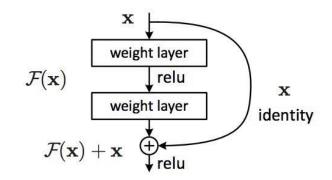

DNN (fully connected)

CNN (weight sharing and convolving)

- Reduce the number of parameters (less overfitting)

- Learn local features

- Translation invariance

- To go **wiiiiiider**

- To extract different features

- To go **deeeeeeper**

- To avoid the **vanishing gradient**

New terms

- Receptive field
- Convolution
- CNN
- Translation invariance
- Stride
- Pooling
- Feature map
- Padding
- Inception
- Skip connection