

ImageNet Classification with Deep Convolutional Neural Networks

by Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton

Kazuki Egusa, 03/12/2019

<https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>

Conclusion

AlexNet : Deep Learning 初の成功例 => 第三次AIブーム到来！！！！

(問題点/背景) ラベル付けされた画像のデータセットでも、比較的小さなものであれば上手く認識できていたが、現実世界に存在する物体は、かなりバラエティ豊かなため、もっと大きな学習用データを使う必要があった。実際、小さな画像データセットの欠点は広く認識されており、最近になって、LabelMeやImageNetなどの、大量の画像のラベル付けされたデータセットを集めることが可能になってきた。

(各種先行研究) これまでもGPU上でCNNを実装する試みは行われてきたが、もちろんAlexNet程ではなかった。

(本研究の意義) ラベル付けされた大量の画像を分類するための深層なCNNを構築する。

Main Purpose

ラベル付けされた大量の画像を分類するための深層なCNNを構築。

How to Verify

ImageNet datasets on ILSVRC-2010 & ILSVRC-2012

Important Points When Compared to Previous Researches

大きなスケールのデータセットでも、ラベル付けされた画像を認識出来るようになった。

Discussions

- ・一つの畳み込み層を抜くだけでエラー率が上がったため、深層であることは重要である。
- ・動画にも応用したい。

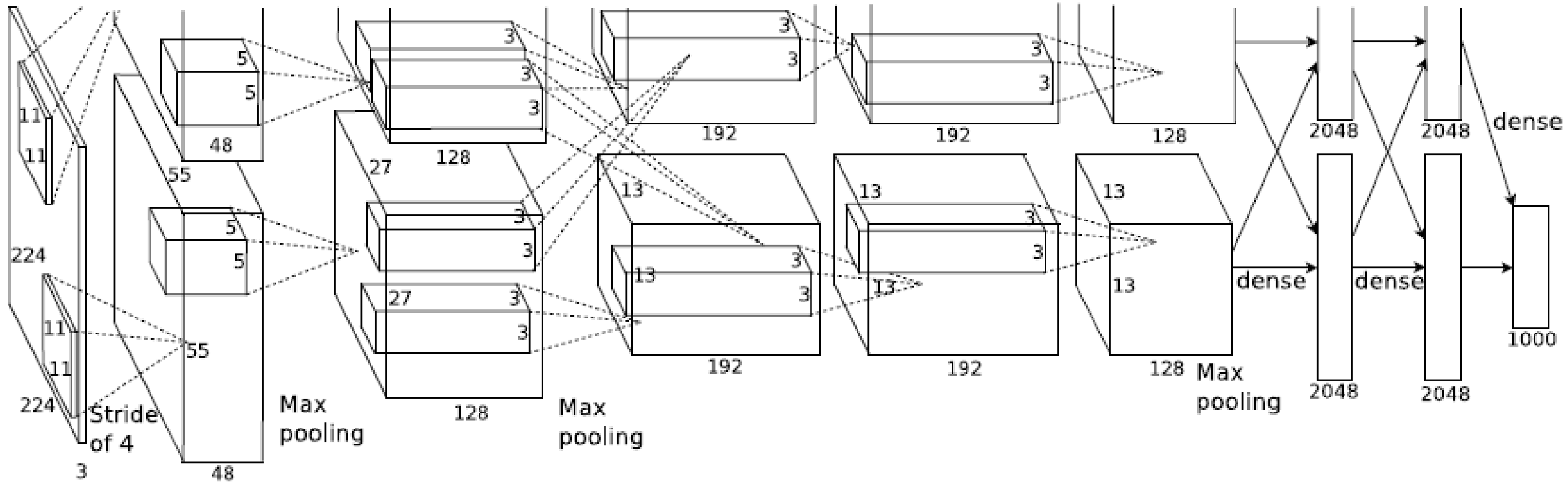
Key Points

Deep learning

Other Treatises I Should Read Next

D.C. Cireşan, U. Meier, J. Masci, L.M. Gambardella, and J. Schmidhuber. High-performance neural networks for visual object classification.
Arxiv preprint arXiv:1102.0183, 2011.

Overall Architectures



5 Convolutional Layers (+ 3 Max Pooling Layers) + 3 Fully Connected Layers + A 1000-way Softmax

Novel or Unusual Features of AlexNet's Architecture (重要度順)

1. 活性化関数: $f(x) = \tanh(x)$ やシグモイド関数 $f(x) = \{1 + e^{(-x)}\}^{-1}$ ではなく、ReLU関数 $f(x) = \max(0, x)$ を採用。
2. 複数のGPUで学習: 今回は二つのGPUを並列化させて使用している。
3. Local Response Normalization (LRN): 全然分かりません。
4. Overlapping Pooling: Pooling層から出力されたunitを、他のunitと被らないようにするのが伝統的だが、条件を満たした場合にのみ被るように取ると、エラー率がわずかに下がり、少しだけ過学習しにくくなった。

Reducing Overfitting

1. Data Augmentation (データの拡張): Label-preserving Transformationを使用して、データセットを人工的に拡張する方法。今回は、画像の形式を変更したものを水平反転させる方法と、学習用データにおいて、RGBチャンネルの強度を変更する方法の二通り。
2. Dropout: 隠れ層の各ニューロンの出力を50%の確率で0にすることで、フォワード/バックプロパゲーションに関与させない(ドロップアウトさせる)。これにより、毎回入力を与えられる度に、ニューラルネットが異なる構築を試すことになるが、全てのニューラルネットで重みは共有される。また、あるニューロンが他のニューロンの出現に依存しなくて済むため、ニューロンの複雑な共適応を減らすことが出来る。ゆえに、よりrobustな特徴量を学習することが出来る。

Results

ILSVRC-2010

Model	Top-1	Top-5
<i>Sparse coding [2]</i>	47.1%	28.2%
<i>SIFT + FVs [24]</i>	45.7%	25.7%
CNN	37.5%	17.0%

ILSVRC-2012

Model	Top-1 (val)	Top-5 (val)	Top-5 (test)
<i>SIFT + FVs [7]</i>	—	—	26.2%
1 CNN	40.7%	18.2%	—
5 CNNs	38.1%	16.4%	16.4%
1 CNN*	39.0%	16.6%	—
7 CNNs*	36.7%	15.4%	15.3%

Results



mite container ship motor scooter leopard

mite	container ship	motor scooter	leopard
black widow	lifeboat	go-kart	jaguar
cockroach	amphibian	moped	cheetah
tick	fireboat	bumper car	snow leopard
starfish	drilling platform	golfcart	Egyptian cat



grille mushroom cherry Madagascar cat

convertible	agaric	dalmatian	squirrel monkey
grille	mushroom	grape	spider monkey
pickup	jelly fungus	elderberry	titi
beach wagon	gill fungus	ffordshire bullterrier	indri
fire engine	dead-man's-fingers	currant	howler monkey

