

Notes de Présentation INF 8225

Olivier Sirois

2018-01-01

Contents

0.1	VGG net	2
0.1.1	Arch	2
0.1.2	Conclusion	2
0.2	Dropout	2
0.2.1	Motivation	2
0.2.2	Principe	3
0.2.3	Implémentation	3
0.2.4	Résultats	3
0.2.5	Autres techniques	3
0.3	Inception	3
0.3.1	Motivation	3
0.3.2	Arch	4
0.3.3	Resultats	4
0.3.4	Conclusion	4
0.4	Batch Norm	4
0.4.1	Motivation	4
0.4.2	résultats	4
0.4.3	Conclusion	5
0.5	ResNet	5
0.5.1	Motivation	5
0.5.2	arch	5
0.5.3	résultats et conclusion	5
0.6	Fully convolutional network for semantic segmentation	5
0.6.1	Motivation et Intuition	5
0.7	Convolutional Netowkrs for Biomedical image segmentation	6
0.7.1	Motivation	6
0.7.2	architecture	6
0.8	YOLO - you only look once, unified real time object detection	6
0.8.1	Motivation + intuition	6
0.8.2	Arch	6
0.8.3	Résultats	7
0.9	Machine Theory of Mind	7
0.9.1	Objectif	7

0.1 VGG net

Réseau de neurones convolutif profond utilisant des petits filtres. On adapte la profondeur à la richesse des données pour éviter la saturation *généralement très profond*.

S'applique à classification + localisation.

0.1.1 Arch

Input images 224x224 RGB preprocessed

Channel propres à chaque couche (64, 128, 256...)

Stride 1

Zero padding Pour préserver la dimension de l'image

Fonction d'activation ReLU

Max Pooling appliqué sur 2x2

FC et softmax pour loss

Différentes configuration sur utiliser pour adapter les données. 16 couche de convolution..

l'objectif est de faire des petits filtre et de les mettre en profondeurs.

0.1.2 Conclusion

Prondeur permet d'améliorer les performance d'un réseau de neurones convolutif au prix de **petits filtres**. Plus performant que GoogleNet lors de son arrivé.

0.2 Dropout

Références, Goodfellow et al. chapitre sur la régularisation...

0.2.1 Motivation

Même motivation que les autres méthodes de régularisation. Réduit l'overfitting et augmente la généralisation sur le model.

0.2.2 Principe

Durant l'entraînement, on applique qu'une sous-partie du réseau déterminé aléatoirement. On 'drop' certain neurones dans le réseau lors de l'entraînement. l'objectif est que le réseau est capable d'apprendre avec l'ensemble de ses neurones au lieu d'avoir certaines neurones qui fait certaines tâches. C'est équivalent à entraîner plusieurs réseaux qui partagent des nodes, puis d'utiliser une moyenne géométrique sur l'assemblage total (bootstrap aggregating??).

0.2.3 Implémentation

On multiplie la sortie de notre neurones avec un masque (1 ou 0) avec une probabilité p .

0.2.4 Résultats

On peut voir une amélioration entre 20 et 60%. Par contre, sa peut prendre beaucoup plus de neurones pour avoir la même performance...

0.2.5 Autres techniques

Augmentation de données, pré-entraînement non-supervisé.. Aucune des techniques ne sont mutuellement exclusive.

0.3 Inception

Architecture de google..

0.3.1 Motivation

Pour améliorer googlenet, essaie de réduire les dimensions des matrice d'entré. Essaie de factoriser les filtre de convolution pour améliorer les performances ?? (whatever that means)..

En factorisation, on obtient de convolution plus petites. On effectue une factorisation spatial dans un filtre asymétrique.

(remplace des convolution de 5x5 par 2 convolution de 3x3. cette opération amène un gains relatif d'environ 28%.

0.3.2 Arch

Plusieurs convolution de taille différente.. (factorisation)

Utilisation de classificateur auxiliaire (en parallèle??)..

Pas compris.. (les gens qui ont présenté n'ont pas trop compris non plus)

0.3.3 Resultats

Plus grande performance dans les réseaux ayant un plus grand champ réceptif.

0.3.4 Conclusion

factorisation des opérations de convolution en convolution plus petite, ce qui permet de rendre model beaucoup plus compact. Ce qui en résulte à une plus grande facilité à mettre à l'échelle des réseaux de grandes tailles.

0.4 Batch Norm

0.4.1 Motivation

La mise à l'échelle de différents features en entrée d'un réseau permet l'apprentissage. On peut l'appliquer au Deep Learning

l'idée est de normaliser les input avant l'activation pour qu'ils aient une distribution précise (moyenne = 0, variance = 1).

0.4.2 résultats

Avec batch norm, on converge rapidement à un model qui apprend. BN rend la distribution de la sortie beaucoup plus stable.

l'application de BN sur inception permet d'améliorer la puissance de l'apprentissage. voir le modèle dans la présentation pour voir la vitesse de convergence. La performance augmente plus rapidement et se stabilise beaucoup mieux (plus performant). Sa convergence beaucoup plus vite car la stabilité de la distribution des sorties est beaucoup plus stable alors on peut se permettre d'augmenter le taux d'apprentissage.

L'utilisation de la normalisation par lot peut nous permettre de nous passer de Dropout pour régulariser. Elle facilite aussi l'utilisation des fonctions sigmoïde dans les réseaux profonds en plus de pouvoir être appliqué aux couches de convolutions.

0.4.3 Conclusion

0.5 ResNet

0.5.1 Motivation

l'utilisation de plusieurs couches en profondeurs permet d'augmenter la capacité en utilisant moins de liens, par contre la profondeur fait en sorte qu'on a souvent un problème de 'vanishing gradient' et un apprentissage très lent.

Pour remédier au problème, On ajoute des connections entre des layers de différentes profondeur pour que le gradient puisse se propager dans les couches plus hautes sans avoir la dégradation causées par toutes les couches lui suivants.

$$H(x) = F(x) + \underline{x}$$

0.5.2 arch

suit les mêmes guideline que le VGG. On applique aussi la normalisation par lot pour normalisé les sorties obtenus.

0.5.3 résultats et conclusion

ResNet à dépassé de loin toutes les architectures de reconnaissance à l'époque. Par contre, elle contenait 157 couches de profondeur comparés aux 15-20 des autres architectures. Cette architecture permet de contourner le problème de disparition de gradient (vanishing gradient).

ResNet s'inspire beaucoup du *highway Network*. ResNet est plus performance que Highway sur CIFAR-10. (6.6% vs. 8.7%).

Certains raffinement en enlevant des fonctions d'activation permet d'augmenter la performance de manière significative.

ResNExt : utilisation des couches en parallèle en plus d'avoir des connections résiduels.

0.6 Fully convolutional network for semantic segmentation

0.6.1 Motivation et Intuition

Des couches fully connected utilise un noyau de la taille de l'image. En ayant un filtre pour chacune des classes à prédire, on peut avoir une carte avec plus

de localités. la classification se fait de manière plus locales.

Une couche fully connected = ??

Trop rapide.. pas toute compris..

0.7 Convolutional Netowkrs for Biomedical image segmentation

0.7.1 Motivation

Utilisation des réseaux de neurones convolutif pour segmenter les iamges médicales. Pour faire les reconnaissance de features pouvant être des précurseur à certaines pathologies.

0.7.2 architecture

Voir photos. on copie l'image d'entrée à la sortie. Sauf qu'on a des branches résiduel qui permet la classification de features contextuel sans affecter l'image originale.

0.8 YOLO - you only look once, unified real time object detection

0.8.1 Motivation + intuition

Provient d'un besoin en temps réel de modèle capable de faire:

- Conduite autonome
- Dispositif d'assistance
- ..

Comparés au approches classique, YOLO fait un seul problème de régression une seule fois dans l'image. Il fait des prédictions directement à partir d'une seule image. 45 FPS pour YOLO

0.8.2 Arch

Utilise seulement un seul réseau convolutif ayant comme input un image complet. Il prédit des boîte délimitant la classification ainsi que sa probabilité. Étape:

- Division de l'image en grille $s \times s$
- Chaque cell prédit un certain nombre de boite de délimitation
- Évaluation d'un taux de confiance
- Évaluation de la probabilité de la classe de l'objet

Fonctions d'activation : ReLU or leaky ReLU

0.8.3 Résultats

Avantages:

- Rapide
- Simple à construire
- Généralisable sur différents domaines
- Assez bonne précision de détection d'object
- Fausse détection relativement faible

Désavantage:

- Loss function peu puissante
- Caractéristique relativement grossière
- Forte contraintes spatiale
- Principale source d'erreur: les localisations

0.9 Machine Theory of Mind

Concept qui est lié à la psychologie. Notre compréhension des autres nous permet d'interagir avec les autres. Le papier propose une manière d'interfacer les ordinateurs avec les humains.

0.9.1 Objectif

Produire un observateur qui peut apprendre autonomement. On essaie d'apprendre à apprendre à modéliser des agents intelligents.

Chapter 1

Semaine 2