Retour sur le TP n°5 (entraînement d’un FCN) + Compléments divers

Pierre Lepetit
ENM, le 12/12

Retour sur le TP 5 :

Au cours des TP n°4/5 nous avons croisé :

- L'opération de normalisation...
  - Codée dans la classe Dataset
  - Cas classiques :
    - Centrage + réduction
    - Normalisation min – max : $y = \dfrac{x - m}{M - m}$

Retour sur le TP 5 :

Au cours des TP n°4/5 nous avons croisé :

- L'opération de normalisation...
- … Qu'il ne faut pas confondre avec les couches de normalisation :

  BatchNorm, InstanceNorm, LayerNorm, GroupNorm, etc

BatchNorm2D :

$$\mu_c = \frac{1}{(BHW)} \sum_{b,h,w} x_{b,c,h,w} \qquad \sigma_c^2 = \frac{1}{(BHW)} \sum_{b,h,w} (x_{b,c,h,w} - \mu_c)^2$$

$$\hat{x}_{b,c,h,w} = \frac{x_{b,c,h,w} - \mu_{b,c}}{\sqrt{\sigma_{b,c}^2 + \epsilon}} \qquad y_{b,c,h,w} = \gamma_c \hat{x}_{b,c,h,w} + \alpha_c$$

En rouge ?

Retour sur le TP 5 :

Au cours des TP n°4/5 nous avons croisé :

- L'opération de normalisation...
- … Qu'il ne faut pas confondre avec les couches de normalisation :

  BatchNorm, InstanceNorm, LayerNorm, GroupNorm, etc

BatchNorm2D :

$$\mu_c = \frac{1}{(BHW)} \sum_{b,h,w} x_{b,c,h,w} \qquad \sigma_c^2 = \frac{1}{(BHW)} \sum_{b,h,w} (x_{b,c,h,w} - \mu_c)^2$$

$$\hat{x}_{b,c,h,w} = \frac{x_{b,c,h,w} - \mu_{b,c}}{\sqrt{\sigma_{b,c}^2 + \epsilon}} \qquad y_{b,c,h,w} = \gamma_c \hat{x}_{b,c,h,w} + \alpha_c$$

Poids entraînables

Retour sur le TP 5 :

Au cours des TP n°4/5 nous avons croisé :

- L'opération de normalisation...
- … Qu'il ne faut pas confondre avec les couches de normalisation :
  BatchNorm, InstanceNorm, LayerNorm, GroupNorm, etc

BatchNorm2D :

$$\mu_c = \frac{1}{(BHW)} \sum_{b,h,w} x_{b,c,h,w} \qquad \sigma_c^2 = \frac{1}{(BHW)} \sum_{b,h,w} (x_{b,c,h,w} - \mu_c)^2$$

$$\hat{x}_{b,c,h,w} = \frac{x_{b,c,h,w} - \mu_{b,c}}{\sqrt{\sigma_{b,c}^2 + \epsilon}} \qquad y_{b,c,h,w} = \gamma_c \hat{x}_{b,c,h,w} + \alpha_c$$

Contexte ?

Retour sur le TP 5 :

Au cours des TP n°4/5 nous avons croisé :

- L'opération de normalisation...
- … Qu'il ne faut pas confondre avec les couches de normalisation :

  BatchNorm, InstanceNorm, LayerNorm, GroupNorm, etc

InstanceNorm2D :

$$\mu_{b,c} = \frac{1}{(HW)} \sum_{h,w} x_{b,c,h,w} \qquad \sigma^2_{b,c} = \frac{1}{(HW)} \sum_{h,w} \left(x_{b,c,h,w} - \mu_{b,c}\right)^2$$

$$\hat{x}_{b,c,h,w} = \frac{x_{b,c,h,w} - \mu_{b,c}}{\sqrt{\sigma^2_{b,c} + \epsilon}} \qquad y_{b,c,h,w} = \gamma_c \hat{x}_{b,c,h,w} + \alpha_c$$

Contexte ?

Retour sur le TP 5 :

Au cours des TP n°4/5 nous avons croisé :

- L'opération de normalisation...
- … Qu'il ne faut pas confondre avec les couches de normalisation :

  BatchNorm, InstanceNorm, LayerNorm, GroupNorm, etc

  LayerNorm :

$$\mu_{b,h} = \frac{1}{K}\sum_{k=1}^{K} x_{b,k,h} \qquad \sigma^2_{b,k} = \frac{1}{H}\sum_{h=1}^{H}\left(x_{b,k,h} - \mu_{b,k}\right)^2$$

$$\hat{x}_{b,k,h} = \frac{x_{b,k,h} - \mu_{b,k}}{\sqrt{\sigma^2_{b,k} + \epsilon}} \qquad y_{b,k,h} = \gamma_h \hat{x}_{b,k,h} + \alpha_h$$

Contexte ?

Retour sur le TP 5 :

Au cours des TP n°4/5 nous avons croisé :

- L'opération de normalisation
- Couches de normalisation
- La convolution « transposée »

Idée 1 : une couche de convolution représente une opération linéaire (si biais nul)

Exemple :

$$\begin{pmatrix} w_{11} & w_{12} & w_{13} & 0 & w_{21} & w_{22} & w_{23} & 0 & w_{31} & w_{32} & w_{33} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{11} & w_{12} & w_{13} & 0 & w_{21} & w_{22} & w_{23} & 0 & w_{31} & w_{32} & w_{33} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{11} & w_{12} & w_{13} & 0 & w_{21} & w_{22} & w_{23} & 0 & w_{31} & w_{32} & w_{33} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{11} & w_{12} & w_{13} & 0 & w_{21} & w_{22} & w_{23} & 0 & w_{31} & w_{32} & w_{33} \end{pmatrix}$$

C'est la matrice d'une conv2D(in_ch=..., out_ch=..., k=..., p=…, s=…, d=...)

de noyau …………….. appliquée à une matrice ... x ...

Retour sur le TP 5 :

Au cours des TP n°4/5 nous avons croisé :

- L’opération de normalisation
- Couches de normalisation
- La convolution « transposée »

Idée 1 : une couche de convolution représente une opération linéaire (si biais nul)

Exemple :

$$\begin{pmatrix} w_{11} & w_{12} & w_{13} & 0 & w_{21} & w_{22} & w_{23} & 0 & w_{31} & w_{32} & w_{33} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_{11} & w_{12} & w_{13} & 0 & w_{21} & w_{22} & w_{23} & 0 & w_{31} & w_{32} & w_{33} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & w_{11} & w_{12} & w_{13} & 0 & w_{21} & w_{22} & w_{23} & 0 & w_{31} & w_{32} & w_{33} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_{11} & w_{12} & w_{13} & 0 & w_{21} & w_{22} & w_{23} & 0 & w_{31} & w_{32} & w_{33} \end{pmatrix}$$

C’est la matrice d’une conv2D(in_ch=1, out_ch=1, k=3, p=0, s=1, d=1)

de noyau $\begin{pmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{pmatrix}$ appliquée à une matrice 4 x 4 : $\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_5 & x_6 & x_7 & x_8 \\ x_9 & x_{10} & x_{11} & x_{12} \\ x_{13} & x_{14} & x_{15} & x_{16} \end{pmatrix}$ vue en colonne

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11} \\ x_{12} \\ x_{13} \\ x_{14} \\ x_{15} \\ x_{16} \end{pmatrix}$$

Retour sur le TP 5 :

Au cours des TP n°4/5 nous avons croisé :

- L'opération de normalisation
- Couches de normalisation
- La convolution « transposée »

Idée 1 : une couche de convolution représente une opération linéaire

Idée 2 : pour passer d'une image de taille 2 x 2 à une image de taille 4 x 4
il suffit d'appliquer la transposée de la matrice précédente.

Cela définit ConvTranspose2D(in_ch=1, out_ch=1, k=3, p=0, s=1, d=1)

La convolution est-elle implémentée comme un produit matriciel ?

Retour sur le TP 5 :

Au cours des TP n°4/5 nous avons croisé :

- L’opération de normalisation
- Couches de normalisation
- La convolution « transposée »
- Le rapport entre minimisation de la MSE et l’espérance conditionnelle

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

METEO FRANCE
À VOS CÔTÉS, DANS UN CLIMAT QUI CHANGE

Retour sur le TP 5 :

Au cours des TP n°4/5 nous avons croisé :

- L’opération de normalisation
- Couches de normalisation
- La convolution « transposée »
- Le rapport entre minimisation de la MSE et l’espérance conditionnelle
- Le rapport entre minimisation de la MAE et la médiane conditionnelle

$$\hat{\theta} = argmin_{\theta} E_{X,Y} L(X,Y) = argmin_{\theta} E_X (E_{Y|X} L(Y,X))$$

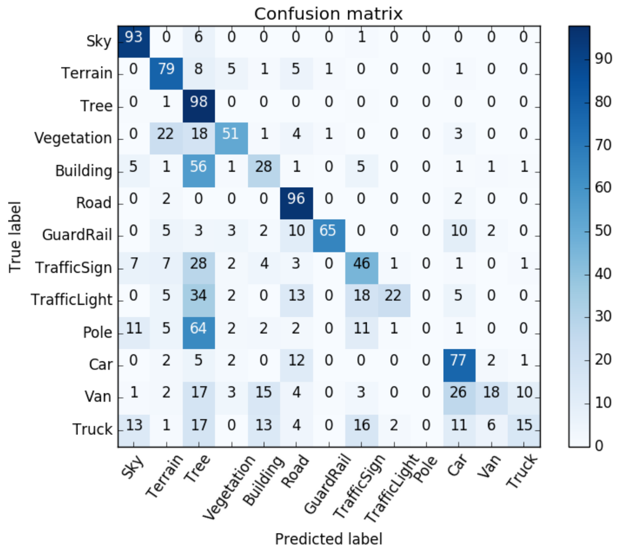
Retour sur le TP 5 :

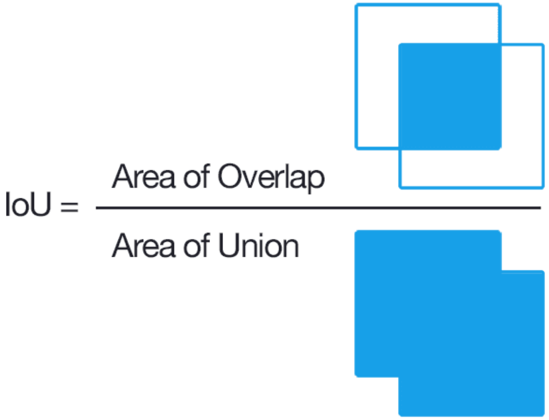
Au cours des TP n°4/5 nous avons croisé :

- L'opération de normalisation
- Couches de normalisation
- La convolution « transposée »
- Le rapport entre minimisation de la MSE et l'espérance conditionnelle
- Le rapport entre minimisation de la MAE et la médiane conditionnelle
- L'utilisation de la mIoU :

Exemple :

IoU(Vegetation) = ?





Retour sur le TP 5 :

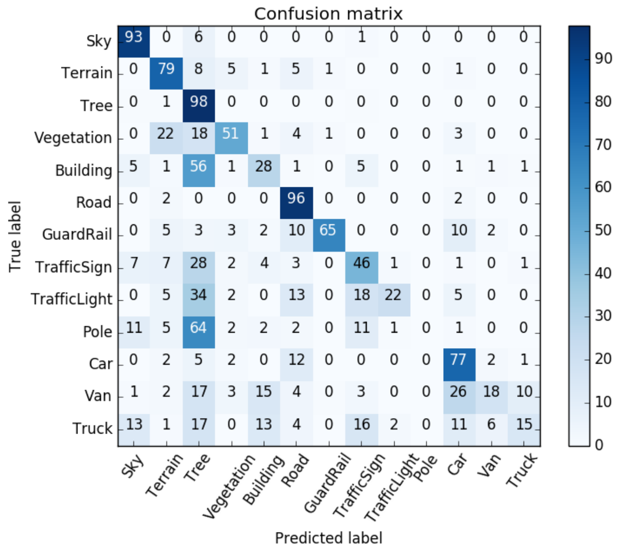
Au cours des TP n°4/5 nous avons croisé :

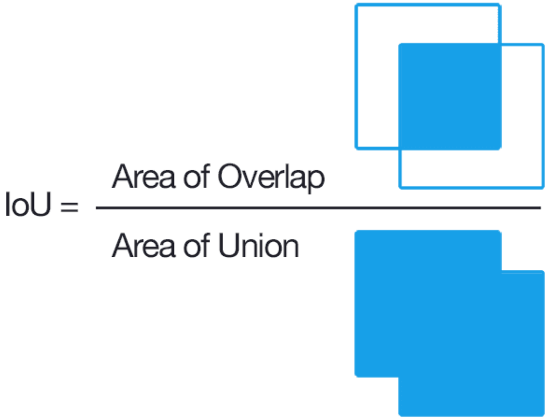
- L'opération de normalisation
- Couches de normalisation
- La convolution « transposée »
- Le rapport entre minimisation de la MSE et l'espérance conditionnelle
- Le rapport entre minimisation de la MAE et la médiane conditionnelle
- L'utilisation de la mIoU :

Exemple :

IoU(Vegetation)
= 51/120

mIoU = 1/13 (IoU(Sky) +
… + IoU(Truck))





Retour sur le TP 5 :

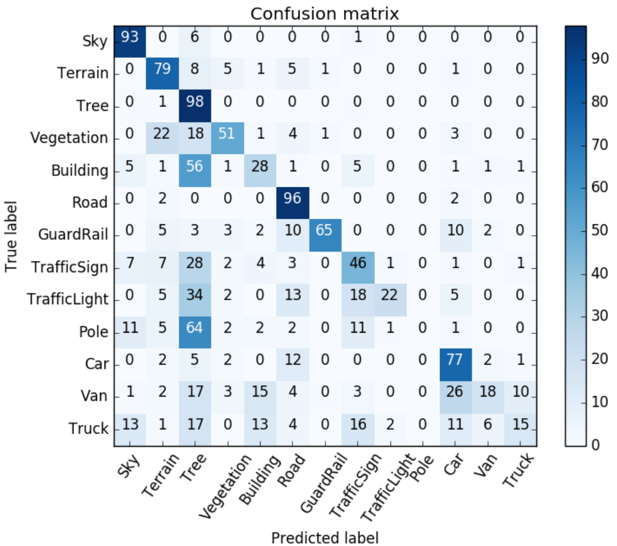
Au cours des TP n°4/5 nous avons croisé :

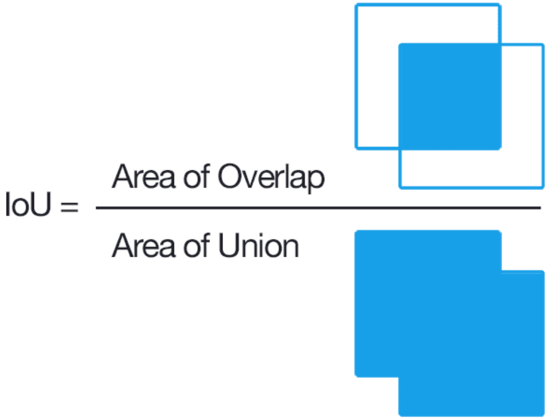
- L'opération de normalisation
- Couches de normalisation
- La convolution « transposée »
- Le rapport entre minimisation de la MSE et l'espérance conditionnelle
- Le rapport entre minimisation de la MAE et la médiane conditionnelle
- L'utilisation de la mIoU :

Exemple :

IoU(Vegetation) = 51/120

mIoU = 1/13 (IoU(Sky) + … + IoU(Truck))





Contexte ?