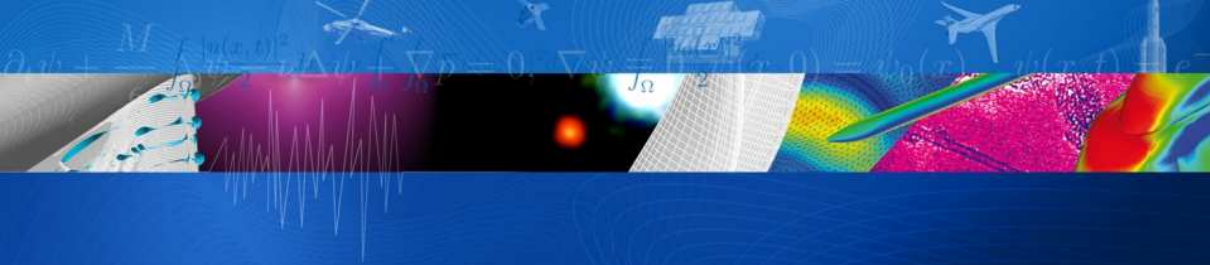


# Apport du flou de défocalisation sur l'estimation de profondeur monoculaire par réseau de neurones

Marcela Carvalho[1,2], Bertrand Le Saux[1], Pauline Trouvé-Peloux[1], Andrés Almansa[2], Frédéric Champagnat[1]

[1]ONERA
[2]Paris Descartes

ONERA
THE FRENCH AEROSPACE LAB

# Estimation de profondeur mono-image


Camera RGB

# Estimation de profondeur mono-image


Camera RGB

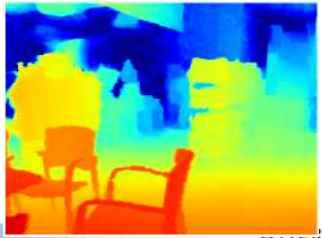
Image RGB

# Estimation de profondeur mono-image


Camera RGB

Image RGB

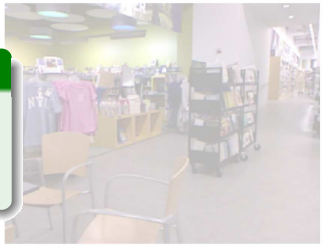
Carte de profondeur



# Estimation de profondeur mono-image

**Avantages**

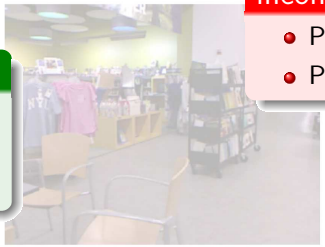
- Compact ;
- Bas coût ;
- Passif.



# Estimation de profondeur mono-image

**Avantages**

- Compact ;
- Bas coût ;
- Passif.



**Inconvénients**

- Pas de correspondance stéréo ;
- Pas de mouvement (vidéo).

# Estimation de profondeur mono-image

## Possibles indices sur les images 2D

- Indices géométriques ;

# Estimation de profondeur mono-image

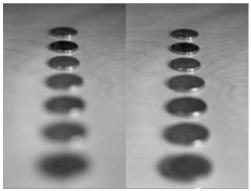
## Possibles indices sur les images 2D

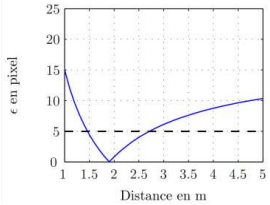
- Indices géométriques ;
- Lignes de fuite ;

# Estimation de profondeur mono-image

## Possibles indices sur les images 2D

- Indices géométriques ;
- Lignes de fuite ;
- Flou de défocalisation.

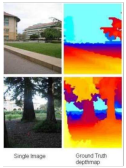


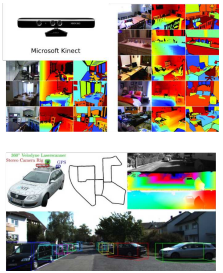


# Estimation de profondeur mono-image

## Bases de données pour l'estimation de la profondeur

- Make3D (Saxena et al., 2009) ;
- NYUv2 (Nathan Silberman & Fergus, 2012) ;
- KITTI (Geiger et al., 2012).

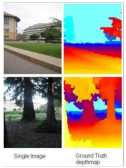


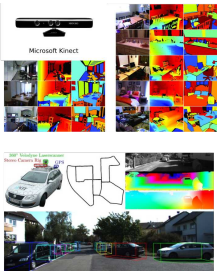



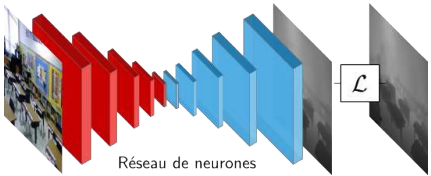
# Estimation de profondeur mono-image

## Bases de données pour l'estimation de la profondeur

- Make3D (Saxena et al., 2009) ;
- NYUv2 (Nathan Silberman & Fergus, 2012) ;
- KITTI (Geiger et al., 2012).

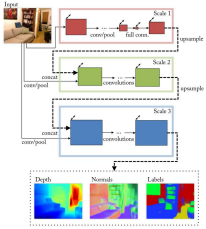



Réseau de neurones convolutif (CNN)

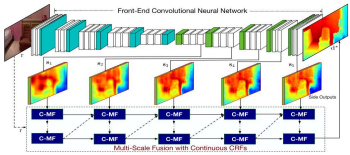
# État de l'art : estimation de profondeur avec les CNNs




(Eigen & Fergus, 2015)

**Caractéristiques**

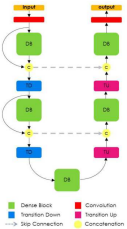
- Architecture multiple-échelle ;
- Fonction de coût invariante à l'échelle.
- $\mathcal{L}_{eigengrad} = \frac{1}{N}\sum_i^N d_i^2 - \frac{\lambda}{2N^2}(\sum_i^N d_i)^2 + \frac{1}{N}\sum_i^N[(\nabla_x di)^2 + (\nabla_y di)^2]$




(Xu et al., 2017)

**Caractéristiques**

- CRF multiple-échelle ;
- Réseau profondément supervisé.
- $\mathcal{L}_2 = \frac{1}{N}\sum_i^N (l_i)^2$

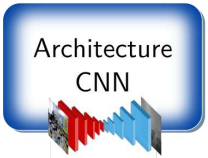



(Jégou et al., 2017 ; Kendall & Gal, 2017)

**Caractéristiques**

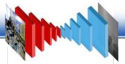
- Connections denses dans l'encodeur et décodeur ;
- Fonctions de coût prennent en compte l'ignorance du modèle.
- $= \frac{1}{N}\sum_i^N \frac{1}{2}exp(-s_i)(l_i)^2 + \frac{1}{2}s_i$

# Sommaire



# Sommaire

Architecture CNN

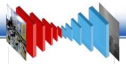


$\mathcal{L}$

Fonctions de coût

# Sommaire

Architecture CNN



$\mathcal{L}$

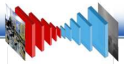
Fonctions de coût

Flou de défocalisation

$\mathcal{E}$

# Sommaire

Architecture CNN



$\mathcal{L}$

Fonctions de coût

Flou de défocalisation

$\mathcal{E}$

Incertitudes du modèle CNN

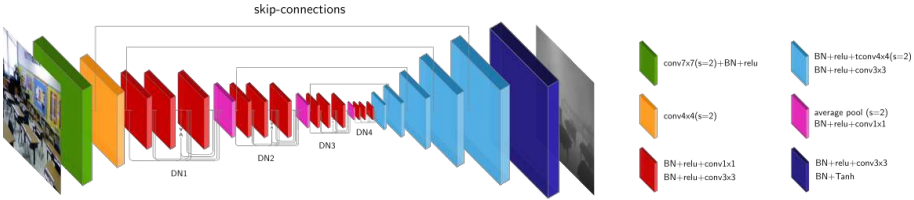
# Le réseau D3-Net

Estimation de profondeur avec des connections denses

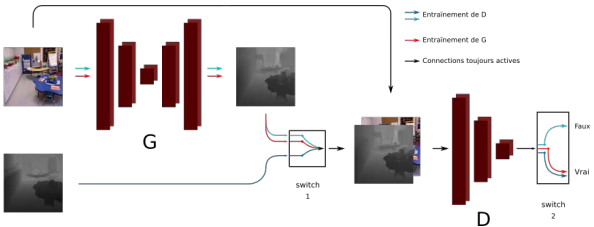
# D3-Net : *Deep Dense Depth estimation Network*

## Architecture proposée

- Exploration des connections denses(Huang et al., 2017) ;
- Exploration des *skip-connections* entre le codeur et le décodeur (Ronneberger et al., 2015).



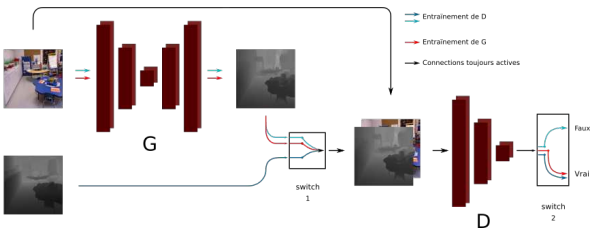
# Le réseau génératif adversaire (GAN)




## L'entraînement adversaire

- Le générateur (G) doit créer des cartes de profondeur pour tromper le discriminateur (D) ;
- Le discriminateur doit être capable de classifier des vrais et faux échantillons.

# Le réseau génératif adversaire (GAN)



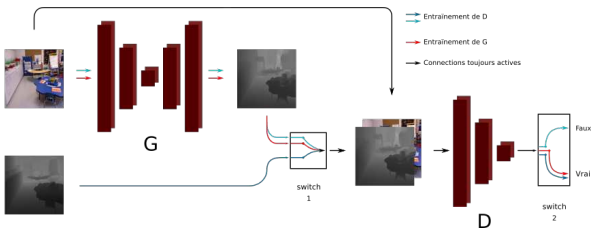

## L'entraînement adversaire

- Le générateur (G) doit créer des cartes de profondeur pour tromper le discriminateur (D) ;
- Le discriminateur doit être capable de classifier des vrais et faux échantillons.

## Avantage

Pas besoin de définir une fonction de coût.

# Le réseau génératif adversaire (GAN)




## L'entraînement adversaire

- Le générateur (G) doit créer des cartes de profondeur pour tromper le discriminateur (D) ;
- Le discriminateur doit être capable de classifier des vrais et faux échantillons.

## Avantage

Pas besoin de définir une fonction de coût.

## Inconvénient

Besoin de beaucoup de données.

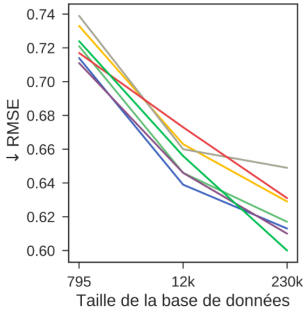
# Étude de fonctions de coût

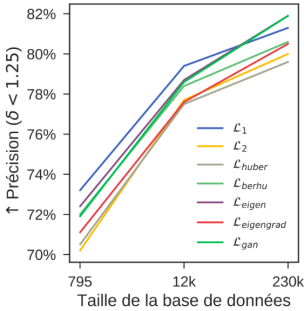
Régression pour l'estimation de la profondeur

# Performance en fonction de la taille de la base de données

**RMSE** $= \sqrt{\frac{1}{N}\sum_{i=0}^{N}(d_i - \hat{d}_i)}$

**Précision** $max(\frac{d_i}{\hat{d}_i}, \frac{\hat{d}_i}{d_i}) = \delta <$

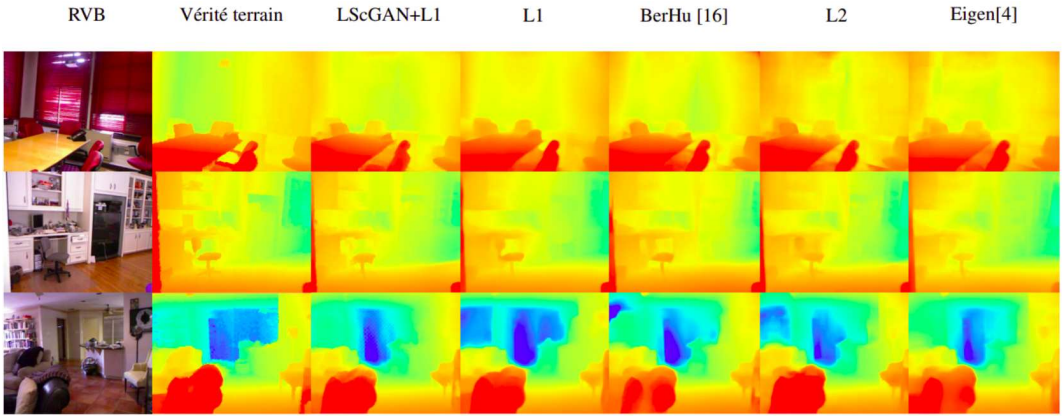



## Considérations

- Plus de données = meilleure performance ;
- Évolution des courbes est différente pour chaque fonction de coût ;
- $\mathcal{L}_1$ et $\mathcal{L}_{eigen}$ ont des bonnes performances en général :
- $\mathcal{L}_{gan}$ bénéficie d'un plus grand nombre de données pour des meilleures prédictions.

# Comparaison qualitative des fonctions de régression



# Comparaison quantitative des fonctions de régression

| Méthodes | Erreur | | | | Précision | | |
|---|---|---|---|---|---|---|---|
| | rel | log10 | rms | rmslog | $\delta < 1.25$ | $\delta < 1.25^2$ | $\delta < 1.25^3$ |
| Images RVB GPC | | | | | | | |
| Saxena [26] | 0.349 | - | 1.214 | - | 44.7% | 74.5% | 89.7% |
| Eigen [3] (VGG16) | 0.158 | - | 0.641 | 0.214 | 76.9% | 95.0% | 98.8% |
| Laina [16] | 0.127 | 0.055 | 0.573 | 0.195 | 81.1% | 95.3% | 98.8% |
| Xu [32] | 0.121 | 0.052 | 0.586 | - | 81.1% | 95.4% | 98.7% |
| Cao [2] | 0.141 | 0.060 | 0.540 | - | 81.9% | 96.5% | 99.2% |
| D3-Net | 0.135 | 0.059 | 0.600 | 0.199 | 81.9% | 95.7% | 98.7% |
| Jung[13] | 0.134 | - | 0.527 | - | 82.2% | 97.1% | 99.3% |
| Kendall and Gal [15] | 0.110 | 0.045 | 0.506 | - | 81.7% | 95.9% | 98.9% |

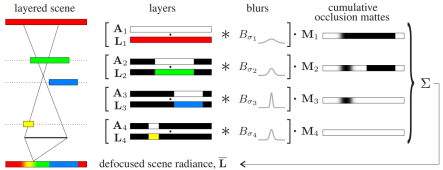
# Le flou comme un indice de profondeur

L'apport du flou de défocalisation pour l'estimation de profondeur

# Génération de la base de données NYUv2 floutée synthétiquement

## Approche par couches successives (Hasinoff & Kutulakos, 2007)

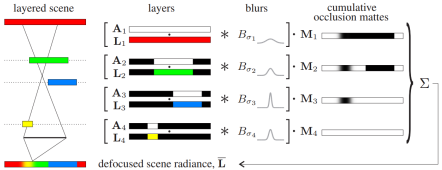
- Somme d'images floutées ;
- Multipliées par des masques en fonction de la profondeur et de l'occlusion des objets en premier plan ;
- Modélisation du flou comme une fonction disque.

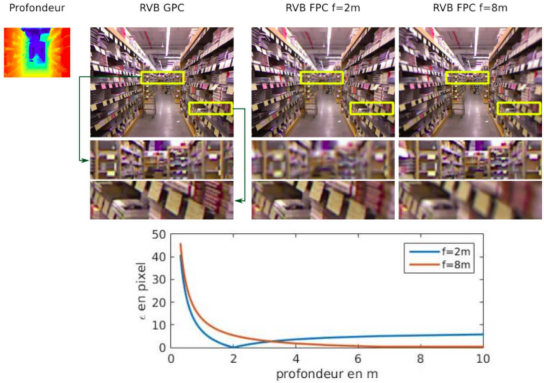


# Génération de la base de données NYUv2 floutée synthétiquement

## Approche par couches successives (Hasinoff & Kutulakos, 2007)

- Somme d'images floutées ;
- Multipliées par des masques en fonction de la profondeur et de l'occlusion des objets en premier plan ;
- Modélisation du flou comme une fonction disque.




Notation :

- GPC : Grande Profondeur de Champs ;
- FPC : Faible Profondeur de Champs.

# Résultats sur l'apport du flou de défocalisation

| Méthodes | Erreur | | | | Précision | | |
|---|---|---|---|---|---|---|---|
| | rel | log10 | rms | rmslog | $\delta < 1.25$ | $\delta < 1.25^2$ | $\delta < 1.25^3$ |
| Images RVB GPC - NYUv2 795 | | | | | | | |
| D3-Net GPC | 0.226 | - | 0.779 | - | 65.8% | 89.2% | 96.7% |
| Images RVB avec flou supplémentaire - NYUv2 795 | | | | | | | |
| D3-Net f=2m | 0.068 | 0.028 | 0.328 | 0.110 | 96.1% | 99.0% | 99.6% |
| D3-Net f=8m | 0.060 | - | 0.403 | - | 95.2% | 99.1% | 99.9% |
| Zhuo *et al.*[33] f=8m) | 0.273 | - | 1.088 | - | 51.7% | 83.1% | 95.1% |
| Trouvé *et al.* [30] f=8m | 0.429 | 0.289 | 1.856 | 0.956 | 39.2% | 52.7% | 61.5% |
| Anwar [1] | 0.094 | 0.039 | 0.347 | - | - | - | - |

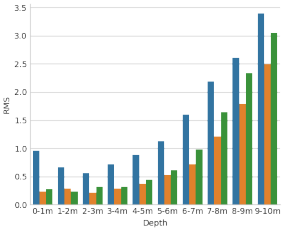
# Résultats sur l'apport du flou de défocalisation

| Méthodes | Erreur | | | | Précision | | |
|---|---|---|---|---|---|---|---|
| | rel | log10 | rms | rmslog | $\delta < 1.25$ | $\delta < 1.25^2$ | $\delta < 1.25^3$ |
| Images RVB GPC - NYUv2 795 | | | | | | | |
| D3-Net GPC | 0.226 | - | 0.779 | - | 65.8% | 89.2% | 96.7% |
| Images RVB avec flou supplémentaire - NYUv2 795 | | | | | | | |
| D3-Net f=2m | 0.068 | 0.028 | 0.328 | 0.110 | 96.1% | 99.0% | 99.6% |
| D3-Net f=8m | 0.060 | - | 0.403 | - | 95.2% | 99.1% | 99.9% |
| Zhuo *et al.*[33] f=8m) | 0.273 | - | 1.088 | - | 51.7% | 83.1% | 95.1% |
| Trouvé *et al.* [30] f=8m | 0.429 | 0.289 | 1.856 | 0.956 | 39.2% | 52.7% | 61.5% |
| Anwar [1] | 0.094 | 0.039 | 0.347 | - | - | - | - |

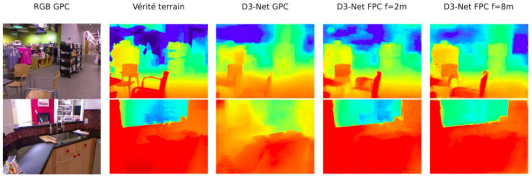
- Amélioration des prédictions ;
- Capacité de surmonter l'ambiguïté du flou de défocalisation ;
- Sensibilité de la performance selon les paramètres.

# Résultats sur l'apport du flou de défocalisation

| Méthodes | Erreur | | | | Précision | | |
|---|---|---|---|---|---|---|---|
| | rel | log10 | rms | rmslog | $\delta < 1.25$ | $\delta < 1.25^2$ | $\delta < 1.25^3$ |
| Images RVB GPC - NYUv2 795 | | | | | | | |
| D3-Net GPC | 0.226 | - | 0.779 | - | 65.8% | 89.2% | 96.7% |
| Images RVB avec flou supplémentaire - NYUv2 795 | | | | | | | |
| D3-Net f=2m | 0.068 | 0.028 | 0.328 | 0.110 | 96.1% | 99.0% | 99.6% |
| D3-Net f=8m | 0.060 | - | 0.403 | - | 95.2% | 99.1% | 99.9% |
| Zhuo *et al.*[33] f=8m) | 0.273 | - | 1.088 | - | 51.7% | 83.1% | 95.1% |
| Trouvé *et al.* [30] f=8m | 0.429 | 0.289 | 1.856 | 0.956 | 39.2% | 52.7% | 61.5% |
| Anwar [1] | 0.094 | 0.039 | 0.347 | - | - | - | - |




- Amélioration des prédictions ;
- Capacité de surmonter l'ambiguïté du flou de défocalisation ;
- Sensibilité de la performance selon les paramètres.

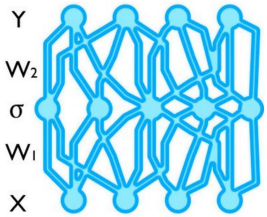


# Étude de l'incertitude du réseau

L'ignorance du modèle par rapport à la distribution à priori

# L'étude de l'incertitude du modèle

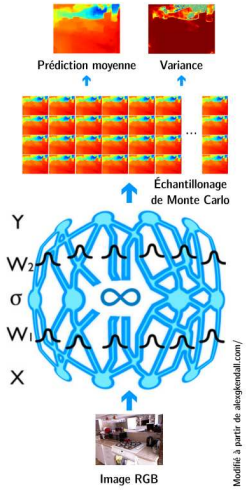
- La connaissance de l'incertitude du réseau nous permet de connaître les limitations du modèle.



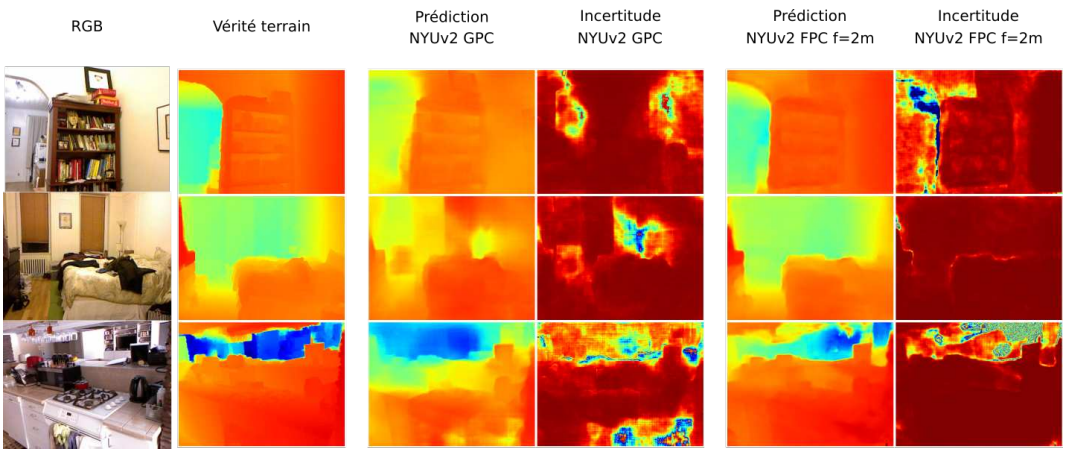
# L'étude de l'incertitude du modèle

La méthode consiste à utiliser un :

- Réseau Bayésien ; et
- la méthode *Monte Carlo dropout* pour générer une carte d'estimation moyenne et de variance.



# Incertitudes épistémiques



# Conclusions

## Conclusions

- $\mathcal{L}_1$ et $\mathcal{L}_{eigen}$ produisent les meilleures performances pour différentes tailles de la base de données ;
- Nous pouvons nous bénéficier d'une fonction de perte adversaire quand nous avons un grand nombre de données ;
- Le flou de défocalisation est un indice important pour l'estimation de la profondeur ;
- Permet d'améliorer les prédictions et réduire l'incertitude du réseau.

# Conclusions

## Conclusions

- $\mathcal{L}_1$ et $\mathcal{L}_{eigen}$ produisent les meilleures performances pour différentes tailles de la base de données ;
- Nous pouvons nous bénéficier d'une fonction de perte adversaire quand nous avons un grand nombre de données ;
- Le flou de défocalisation est un indice important pour l'estimation de la profondeur ;
- Permet d'améliorer les prédictions et réduire l'incertitude du réseau.

## Inconvénients

- Il n'existe pas des bases de données floutées réelles.

# Conclusions

## Conclusions

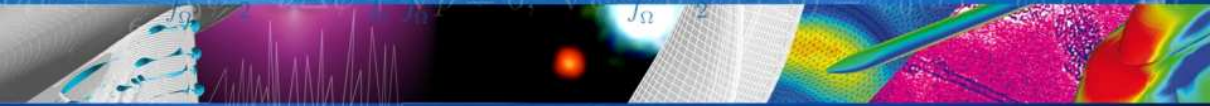
- $\mathcal{L}_1$ et $\mathcal{L}_{eigen}$ produisent les meilleures performances pour différentes tailles de la base de données ;
- Nous pouvons nous bénéficier d'une fonction de perte adversaire quand nous avons un grand nombre de données ;
- Le flou de défocalisation est un indice important pour l'estimation de la profondeur ;
- Permet d'améliorer les prédictions et réduire l'incertitude du réseau.

## Inconvénients

- Il n'existe pas des bases de données floutées réelles.

## Perspectives

- Création d'une base de données avec un capteur DSLR, Kinect, stereo.



# Merci !

marcela.carvalho@onera.fr

ONERA
THE FRENCH AEROSPACE LAB

# Bibliographie succinte


Eigen, D., & Fergus, R. (2015). Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture. *ICCV*.

Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving ? the kitti vision benchmark suite. In *Computer vision and pattern recognition (cvpr), 2012 ieee conference on* (pp. 3354–3361).

Hasinoff, S. W., & Kutulakos, K. N. (2007, Oct). A layer-based restoration framework for variable-aperture photography. In *2007 ieee 11th international conference on computer vision* (p. 1-8). doi: 10.1109/ICCV.2007.4408898

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Cvpr*.

Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017). The one hundred layers tiramisu : Fully convolutional densenets for semantic segmentation. In *Cvprw* (pp. 1175–1183).

Kendall, A., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision ? *arXiv preprint arXiv :1703.04977*.

Nathan Silberman, P. K., Derek Hoiem, & Fergus, R. (2012). Indoor segmentation and support inference from rgbd images. In *Eccv*.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net : Convolutional networks for