

Apport du flou de défocalisation pour l'estimation de profondeur monoculaire par réseaux de neurones

Marcela Carvalho¹, Bertrand Le Saux¹, Pauline Trouvé-Peloux¹,
Andrés Almansa², Frédéric Champagnat¹

¹DTIS, ONERA, Université Paris Saclay
²Université Paris Descartes

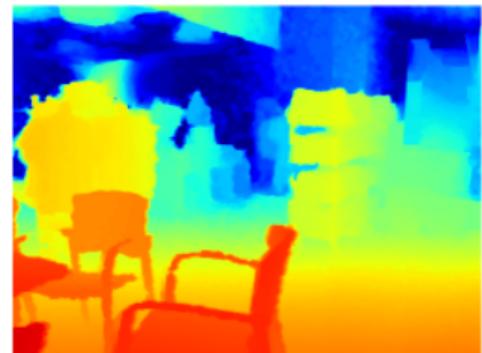
Estimation de profondeur mono-image



Caméra RGB



Image RGB



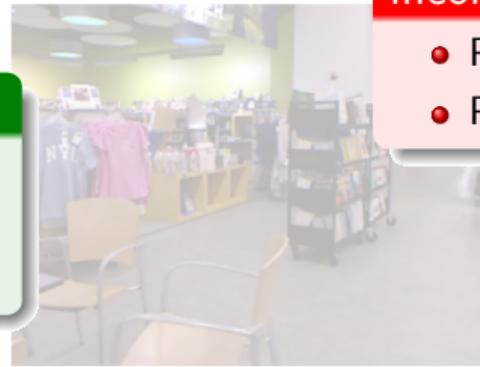
Carte de profondeur

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

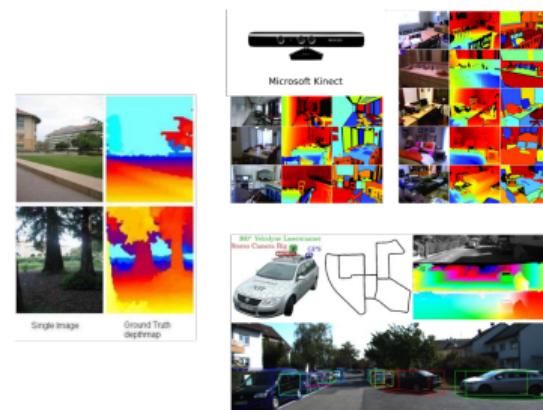
- Indices de contenu ;
- Lignes de fuite ;
- Flou de défocalisation.



Estimation de profondeur mono-image

Bases de données pour l'estimation de la profondeur

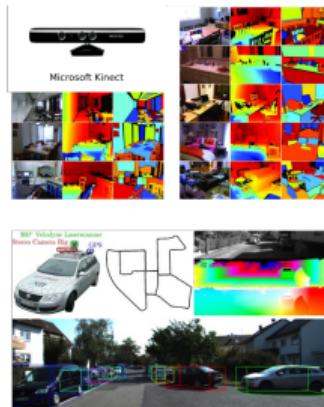
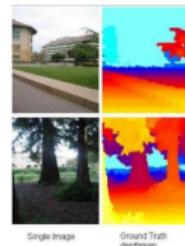
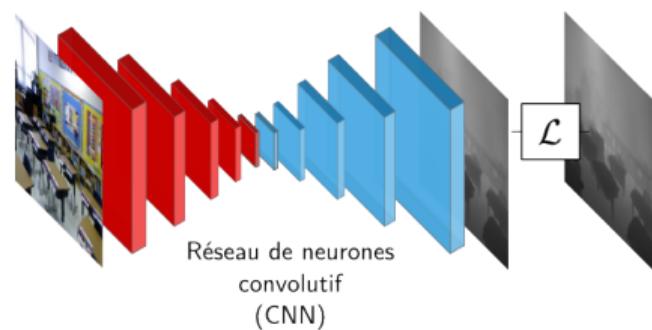
- Make3D (Saxena et al., 2009) ;
- NYUv2 (Nathan Silberman & Fergus, 2012) ;
- KITTI (Geiger et al., 2012).



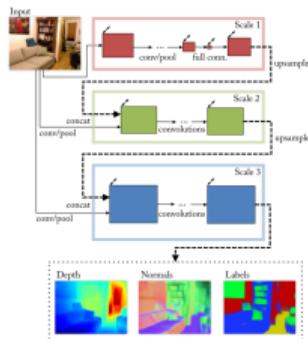
Estimation de profondeur mono-image

Bases de données pour l'estimation de la profondeur

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



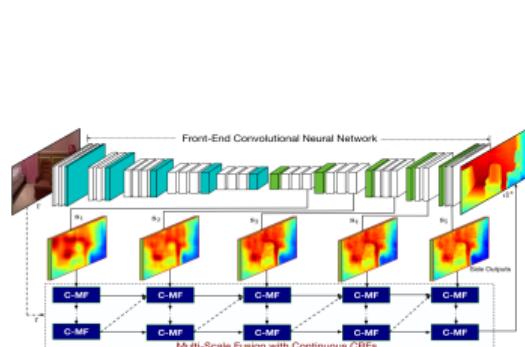
État de l'art : estimation de profondeur avec les CNNs



(Eigen & Fergus, 2015)

Caractéristiques

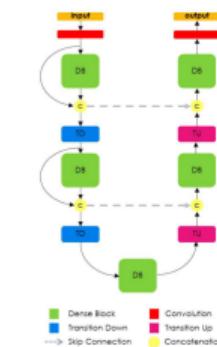
- Architecture multiple-échelle ;
- Fonction de coût invariante à l'échelle.
- $\mathcal{L}_{eigengrad}$
 $= \frac{1}{N} \sum_i^N d_i^2 - \frac{\lambda}{2N^2} (\sum_i^N d_i)^2 + \frac{1}{N} \sum_i^N [(\nabla_x d_i)^2 + (\nabla_y d_i)^2]$



(Xu et al., 2017)

Caractéristiques

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



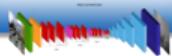
(Jégou et al., 2017; Kendall & Gal, 2017)

Caractéristiques

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

Sommaire

Architecture
CNN



\mathcal{L}
Fonctions
de coût

Flou de
défocalisation



Incertitudes
du modèle CNN



Données
Réelles



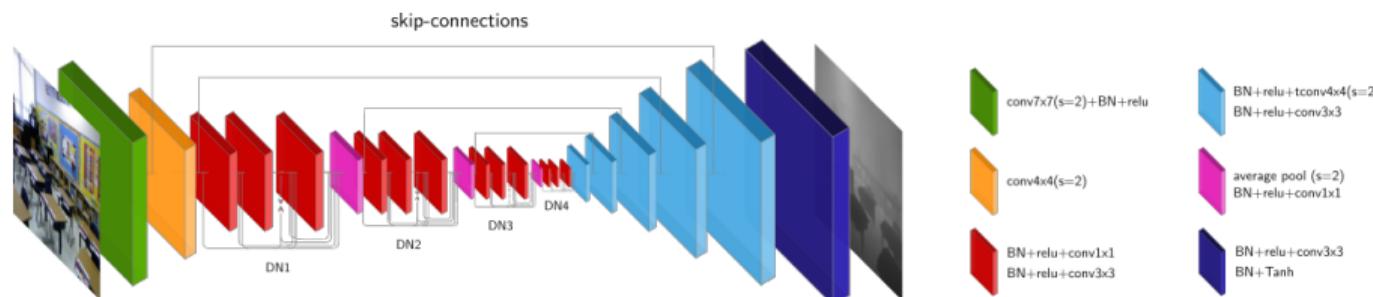
Le réseau D3-Net

Estimation de profondeur avec des connections denses

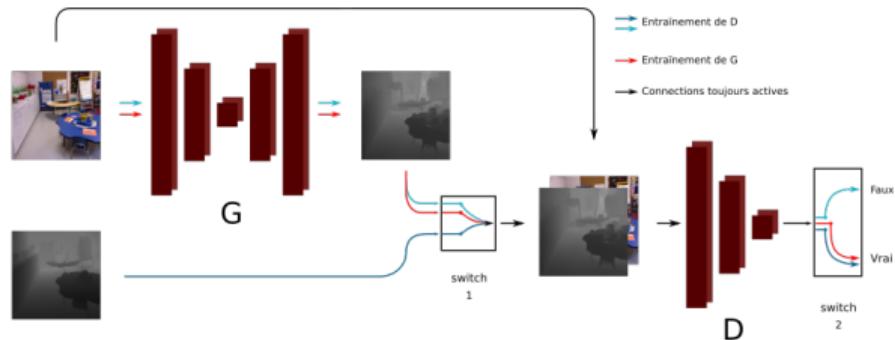
D3-Net : Deep Dense Depth estimation Network

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.

Avantage

Apprentissage de la fonction de coût.

Inconvénient

Entraînement très instable.

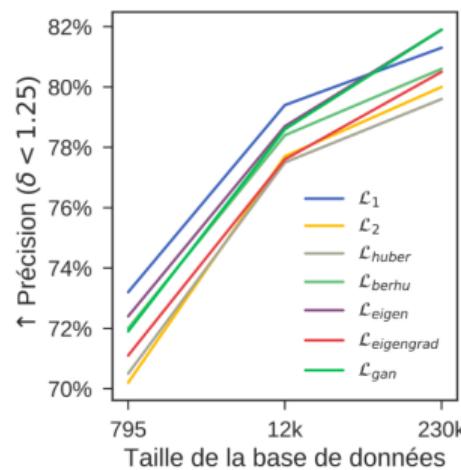
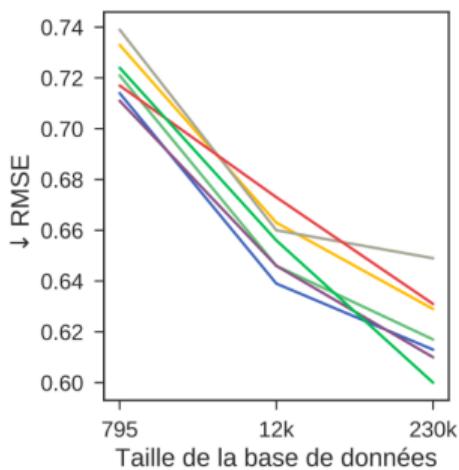
Étude de fonctions de coût

Régression pour l'estimation de la profondeur

Performance en fonction de la taille de la base de données

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=0}^N (d_i - \hat{d}_i)^2}$$

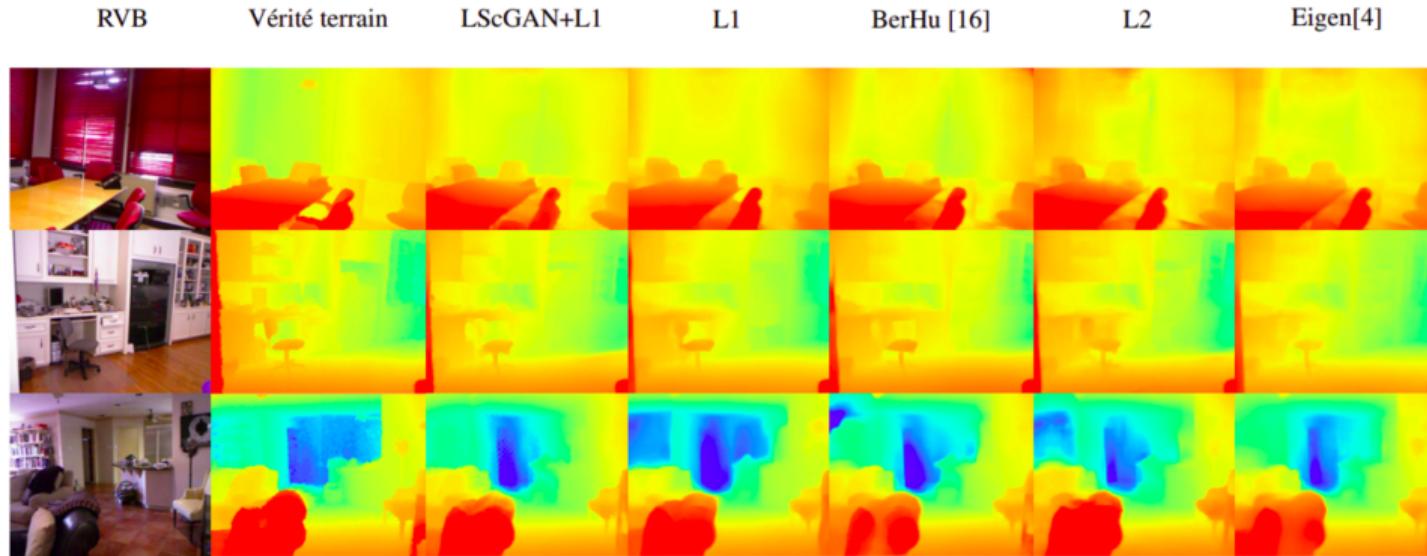
$$\text{Précision } \max\left(\frac{d_i}{\hat{d}_i}, \frac{\hat{d}_i}{d_i}\right) = \delta < \text{thr}$$



Considérations

- Plus de données = meilleure performance ;
- Évolution des courbes est différente pour chaque fonction de coût ;
- \mathcal{L}_1 et $\mathcal{L}_{\text{eigen}}$ 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



→ Carvalho et al., *On Regression Losses for Deep Depth Estimation*, ICIP 2018

Comparaison quantitative des fonctions de régression

Methods	Error↓				Accuracy↑		
	rel	log10	rms	rmslog	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Eigen <i>et al.</i>	0.158	-	0.641	0.214	76.9%	95.0%	98.8%
Laina <i>et al.</i>	0.127	0.055	0.573	0.195	81.1%	95.3%	98.8%
Xu <i>et al.</i>	0.121	0.052	0.586	-	81.1%	95.4%	98.7%
Jung <i>et al.</i>	0.134	-	0.527	-	82.2%	97.1%	99.3%
Kendall and Gal <i>et al.</i>	0.110	0.045	0.506	-	81.7%	95.9%	98.9%
D3-Net*	0.136	-	0.504	-	82.1%	95.5%	98.7%
DORN	0.115	0.051	0.509	-	82.8%	96.5%	99.2%

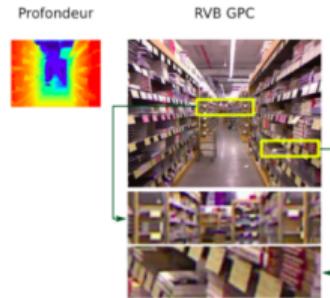
- D3-Net est parmi les meilleures méthodes de l'état de l'art ;
- L'entraînement est fait en une seule phase ;
- Pas besoin d'étape de raffinement.

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

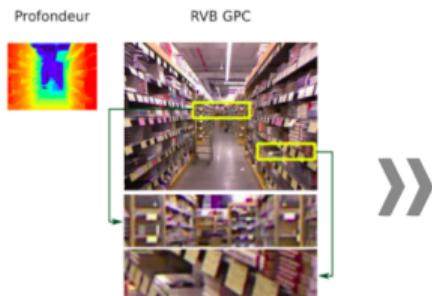
Image nette et carte de profondeur



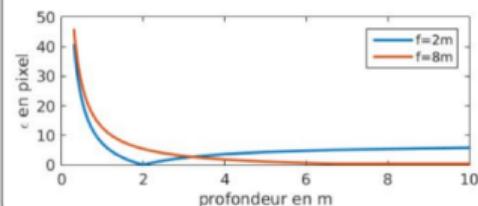
GPC: grande profondeur de champs

Génération de la base de données NYUv2 floutée synthétiquement

Image nette et carte de profondeur

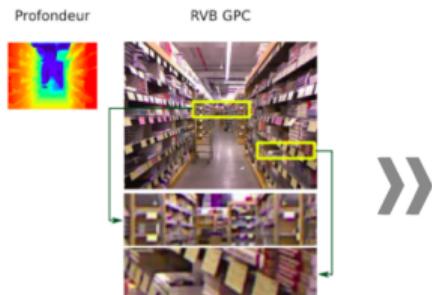


Génération du flou par une approche de couches successives (Hasinoff and Kutulakos, 2007)



Génération de la base de données NYUv2 floutée synthétiquement

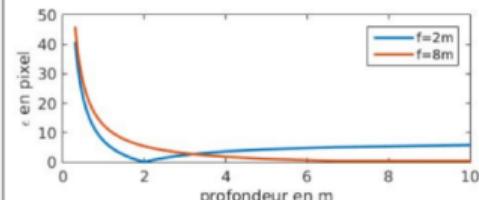
Image nette et carte de profondeur



FPC: faible profondeur de champs

GPC: grande profondeur de champs

Génération du flou par une approche de couches successives
(Hasinoff and Kutulakos, 2007)

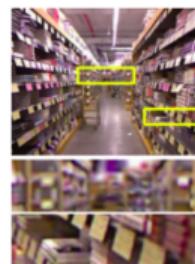


Images floutées

RVB FPC $f=8m$



RVB FPC $f=2m$



Résultats sur l'apport du flou de défocalisation

Methods	Error				Accuracy		
	rel	log10	rms	rmslog	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Original RGB images							
D3-Net All-in-focus	0.226	-	0.779	-	65.8%	89.2%	96.7%
RGB images with additional blur							
D3-Net 2m focus	0.068	0.028	0.328	0.110	96.1%	99.0%	99.6%
D3-Net 4m focus	0.085	0.036	0.465	0.125	92.5%	99.0%	99.8%
D3-Net 8m focus	0.060	-	0.403	-	95.2%	99.1%	99.9%
Zhuo <i>et al.</i> [41] 8m focus)	0.273	-	1.088	-	51.7%	83.1%	95.1%
Trouvé <i>et al.</i> [35] 8m focus	0.429	0.289	1.856	0.956	39.2%	52.7%	61.5%
RGB images with additional blur proposed by [1]							
Anwar [1]	0.094	0.039	0.347	-	-	-	-
D3-Net	0.036	0.016	0.171	0.054	99.3%	100.0%	100.0%

FPC: faible profondeur de champ

GPC: grande profondeur de champ

- Entraînement par *patches* ;
- Fonction de coût \mathcal{L}_1 ;

Résultats sur l'apport du flou de défocalisation

Methods	Error				Accuracy		
	rel	log10	rms	rmslog	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Original RGB images							
D3-Net All-in-focus	0.226	-	0.779	-	65.8%	89.2%	96.7%
RGB images with additional blur							
D3-Net 2m focus	0.068	0.028	0.328	0.110	96.1%	99.0%	99.6%
D3-Net 4m focus	0.085	0.036	0.465	0.125	92.5%	99.0%	99.8%
D3-Net 8m focus	0.060	-	0.403	-	95.2%	99.1%	99.9%
Zhuo <i>et al.</i> [41] 8m focus)	0.273	-	1.088	-	51.7%	83.1%	95.1%
Trouvé <i>et al.</i> [35] 8m focus	0.429	0.289	1.856	0.956	39.2%	52.7%	61.5%
RGB images with additional blur proposed by [1]							
Anwar [1]	0.094	0.039	0.347	-	-	-	-
D3-Net	0.036	0.016	0.171	0.054	99.3%	100.0%	100.0%

FPC: faible profondeur de champ

GPC: grande profondeur de champ

● Amélioration des prédictions ;

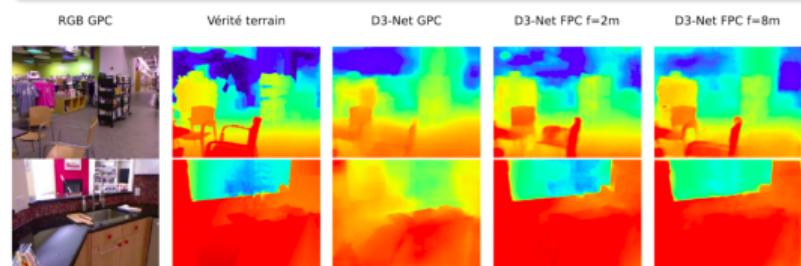
Résultats sur l'apport du flou de défocalisation

Methods	Error				Accuracy		
	rel	log10	rms	rmslog	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Original RGB images							
D3-Net All-in-focus	0.226	-	0.779	-	65.8%	89.2%	96.7%
RGB images with additional blur							
D3-Net 2m focus	0.068	0.028	0.328	0.110	96.1%	99.0%	99.6%
D3-Net 4m focus	0.085	0.036	0.465	0.125	92.5%	99.0%	99.8%
D3-Net 8m focus	0.060	-	0.403	-	95.2%	99.1%	99.9%
Zhuo <i>et al.</i> [41] 8m focus)	0.273	-	1.088	-	51.7%	83.1%	95.1%
Trouvé <i>et al.</i> [35] 8m focus	0.429	0.289	1.856	0.956	39.2%	52.7%	61.5%
RGB images with additional blur proposed by [1]							
Anwar [1]	0.094	0.039	0.347	-	-	-	-
D3-Net	0.036	0.016	0.171	0.054	99.3%	100.0%	100.0%

FPC: faible profondeur de champ

GPC: grande profondeur de champ

- Sensibilité de la performance selon les paramètres ;



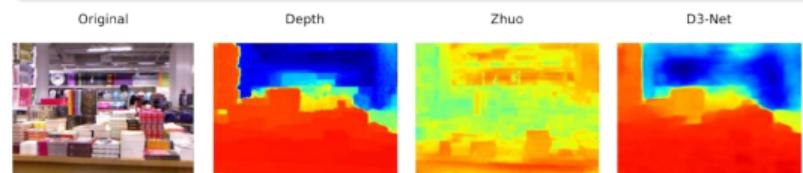
Résultats sur l'apport du flou de défocalisation

Methods	Error				Accuracy		
	rel	log10	rms	rmslog	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Original RGB images							
D3-Net All-in-focus	0.226	-	0.779	-	65.8%	89.2%	96.7%
RGB images with additional blur							
D3-Net 2m focus	0.068	0.028	0.328	0.110	96.1%	99.0%	99.6%
D3-Net 4m focus	0.085	0.036	0.465	0.125	92.5%	99.0%	99.8%
D3-Net 8m focus	0.060	-	0.403	-	95.2%	99.1%	99.9%
Zhuo <i>et al.</i> [41] 8m focus)	0.273	-	1.088	-	51.7%	83.1%	95.1%
Trouvé <i>et al.</i> [35] 8m focus	0.429	0.289	1.856	0.956	39.2%	52.7%	61.5%
RGB images with additional blur proposed by [1]							
Anwar [1]	0.094	0.039	0.347	-	-	-	-
D3-Net	0.036	0.016	0.171	0.054	99.3%	100.0%	100.0%

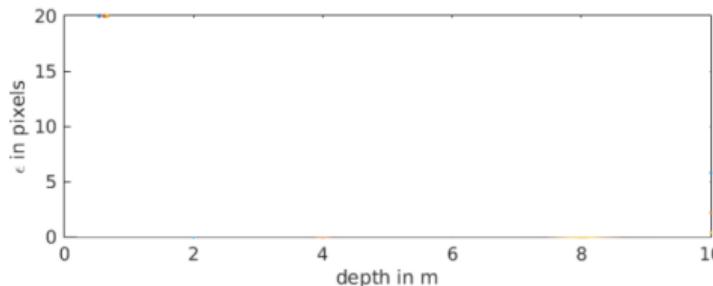
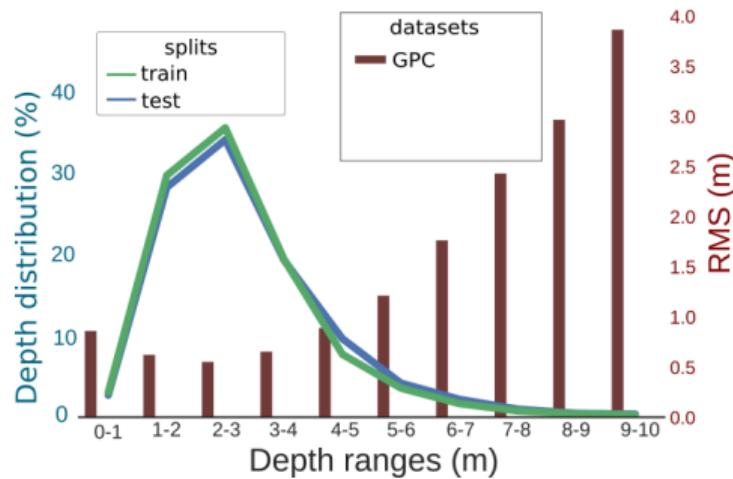
FPC: faible profondeur de champ

GPC: grande profondeur de champ

- Capacité de surmonter l'ambiguïté du flou de défocalisation (Zhuo 2m focus).

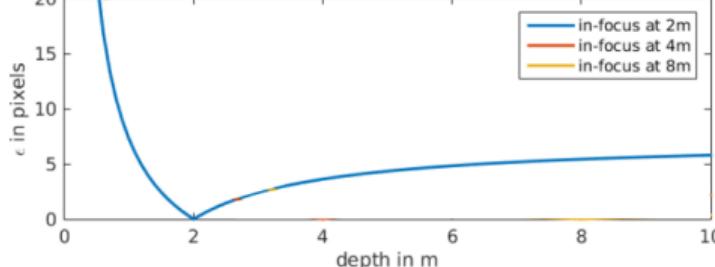
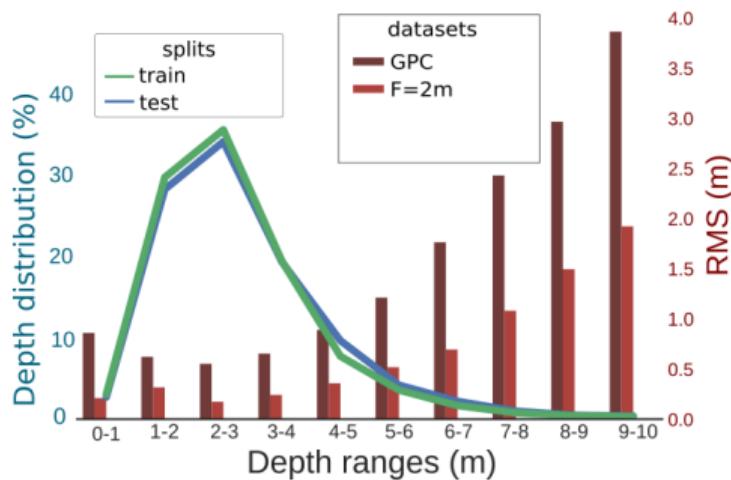


Résultats sur l'apport du flou de défocalisation



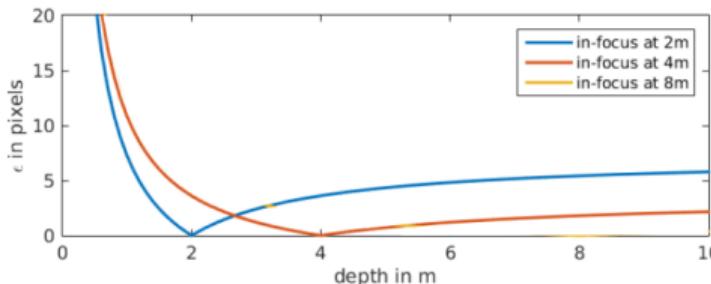
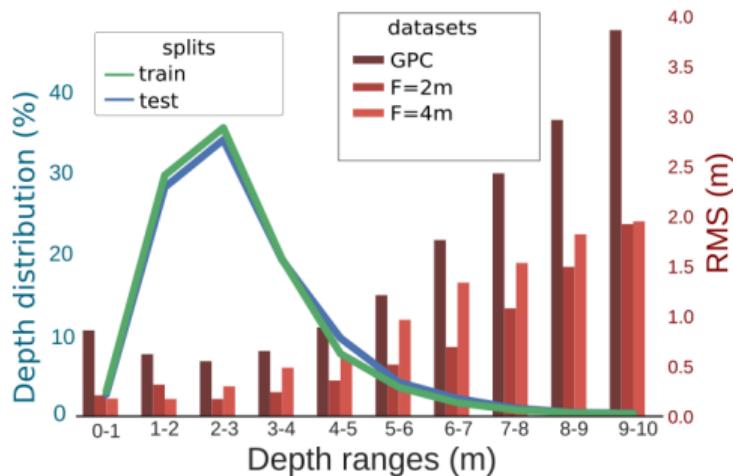
- ↑ données = ↓ erreur ;

Résultats sur l'apport du flou de défocalisation



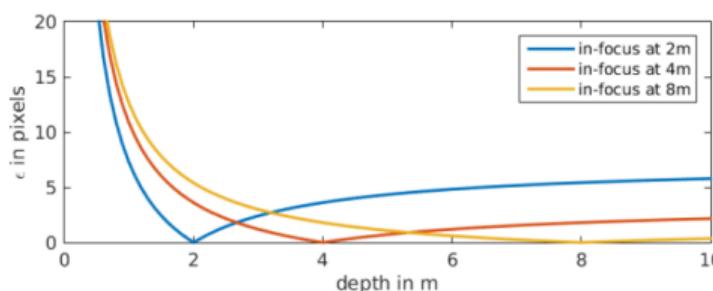
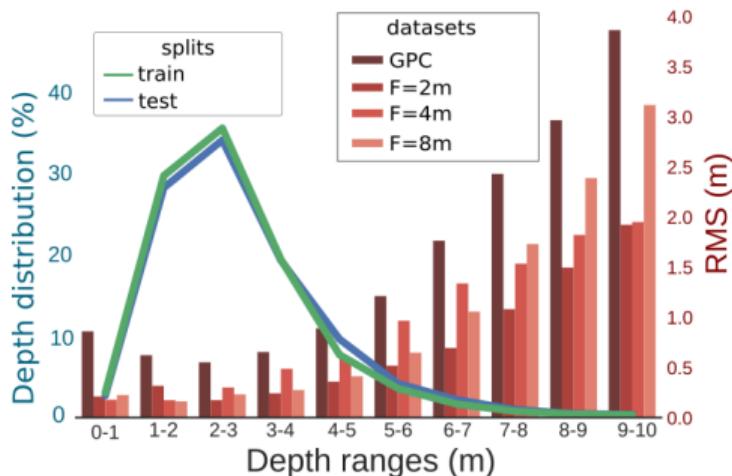
- ↑ données = ↓ erreur ;
- Le flou améliore la performance pour objets peu fréquents ;

Résultats sur l'apport du flou de défocalisation



- ↑ données = ↓ erreur ;
- Le flou améliore la performance pour objets peu fréquents ;
- La RMS évolue différemment suivant le réglage de la caméra ;

Résultats sur l'apport du flou de défocalisation



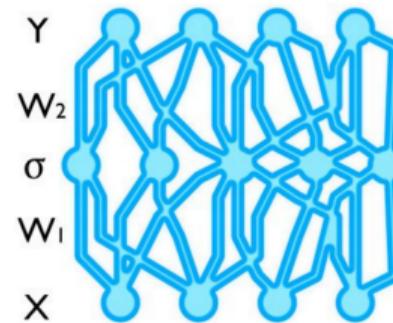
- ↑ données = ↓ erreur ;
- Le flou améliore la performance pour objets peu fréquents ;
- La RMS évolue différemment suivant le réglage de la caméra ;
- Quel est le réglage optimal ?

Étude de l'incertitude du réseau

Comment mesurer la confiance dans les prédictions d'un modèle ?

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 (Kendall et al., 2015) consiste à utiliser :

- un Réseau avec drop-out Bayésien ; et

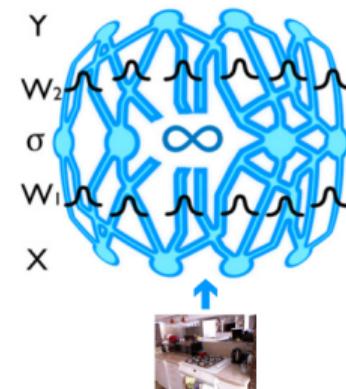


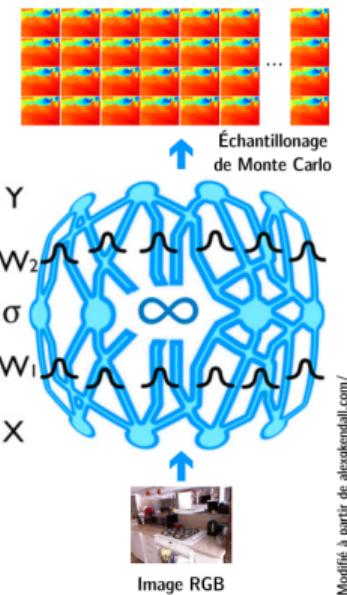
Image RGB

Modifié à partir de alexkendall.com/

L'étude de l'incertitude du modèle

La méthode (Kendall et al., 2015) consiste à utiliser :

- Réseau avec drop-out Bayésien ; et
- une approche Monte Carlo dropout pour générer une carte d'estimation moyenne et de variance.

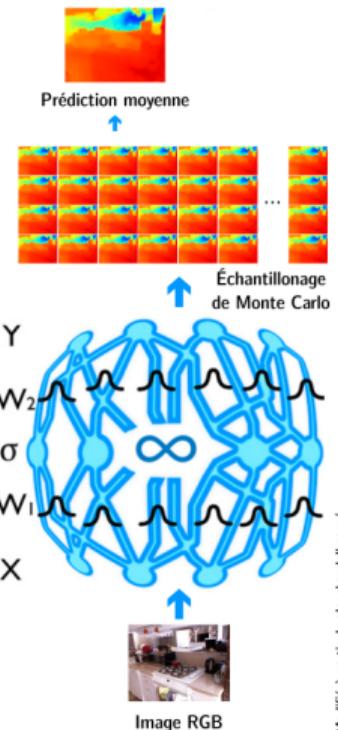


Modifié à partir de alexgkendall.com/

L'étude de l'incertitude du modèle

La méthode (Kendall et al., 2015) consiste à utiliser :

- Réseau avec drop-out Bayésien ; et
- une approche *Monte Carlo dropout* pour générer une carte d'estimation moyenne et de variance.

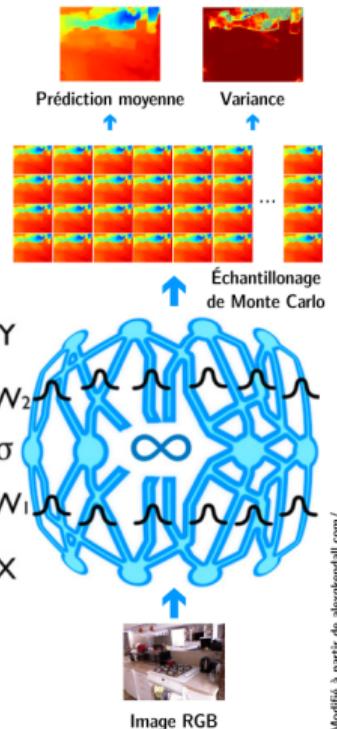


Modifié à partir de alexkendall.com/

L'étude de l'incertitude du modèle

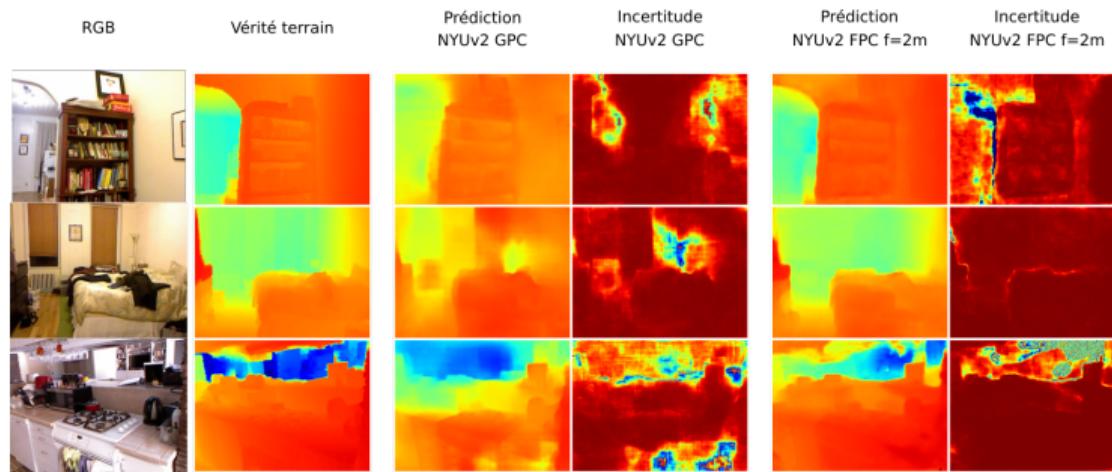
La méthode (Kendall et al., 2015) consiste à utiliser :

- Réseau avec drop-out Bayésien ; et
- une approche *Monte Carlo dropout* pour générer une carte d'estimation moyenne et de variance.



Modifié à partir de alexkendall.com/

L'incertitude du modèle et le flou



- Incertitudes sur zones avec peu de texture ;
- Avec le flou : les erreurs \downarrow et la confiance du réseau \uparrow ;
- Les indices géométriques \downarrow l'ambiguïté du flou.

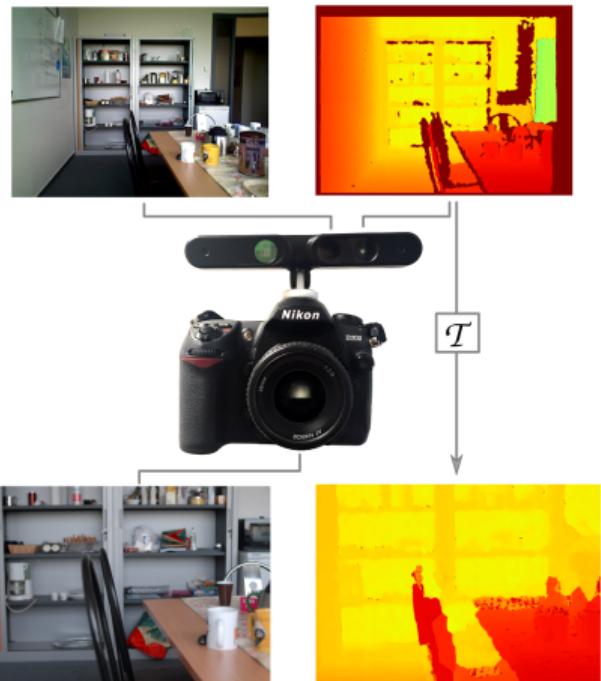
☞ Carvalho et al., *Estimation de profondeur monoculaire par réseau de neurones*, RFIAP 2018

Deep DFD avec des données réelles

Transfert de domaine d'apprentissage

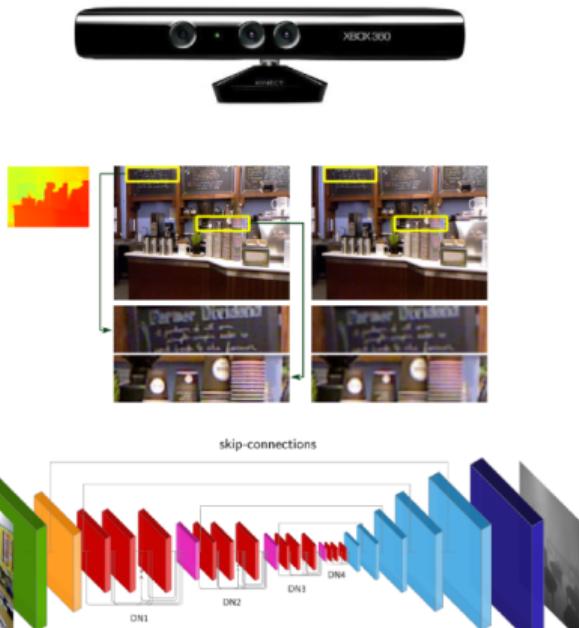
Deep DFD avec des données réelles

- Création d'une base de données réelles avec un reflex et un capteur RGB-D ;



Deep DFD avec des données réelles

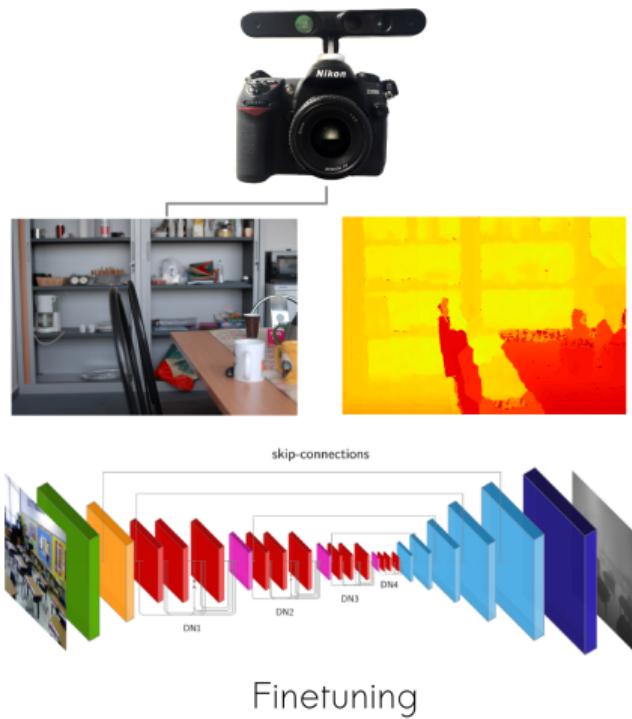
- Création d'une base de données réelles avec un reflex et un capteur RGB-D ;
- Pré-entraînement sur NYUv2 floutée pour simuler le flou de la caméra ;



Pré-entraînement

Deep DFD avec des données réelles

- Création d'une base de données réelles avec un reflex et un capteur RGB-D ;
- Pré-entraînement sur NYUv2 floutée pour simuler le flou de la caméra ;
- *Finetuning* sur la base de données réelles.



Résultats Deep DFD avec des données réelles

Methods	Error↓				Accuracy↑		
	rel	log10	rms	rmslog	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
N=2.8	0.157	0.065	0.592	0.234	80.9%	94.4%	97.6%
N=8	0.225	0.095	0.770	0.285	60.2%	87.7%	98.0%
N=8 (resize)	0.199	0.084	0.695	0.259	69.6%	91.6%	97.4%

FPC: faible profondeur de champ

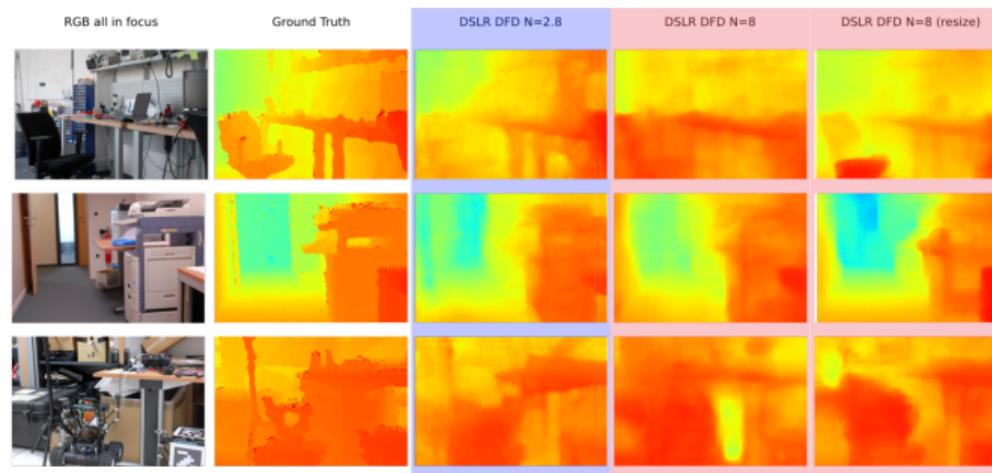
GPC: grande profondeur de champ

FPC
GPC

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

- ↑ performance sur toutes les métriques ;

Résultats Deep DFD avec des données réelles



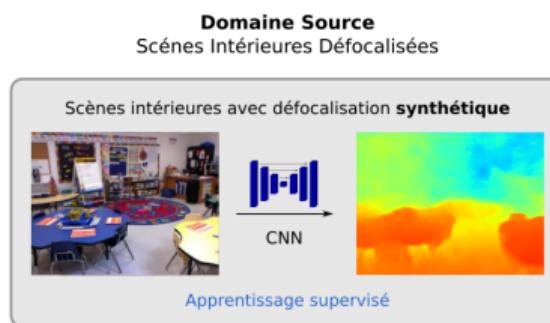
FPC: faible profondeur de champ GPC: grande profondeur de champ

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

- ↑ performance sur toutes les métriques ;
- ↑ segmentation des objets ;
- ↓ erreurs sur objets peu fréquents pendant l'entraînement.

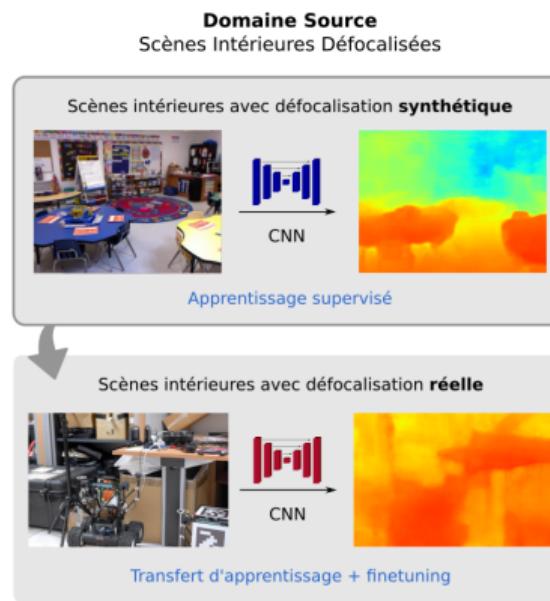
Deep DFD 'in the wild'

Objectif : vérifier la capacité de généralisation du modèle CNN avec des scènes extérieures.



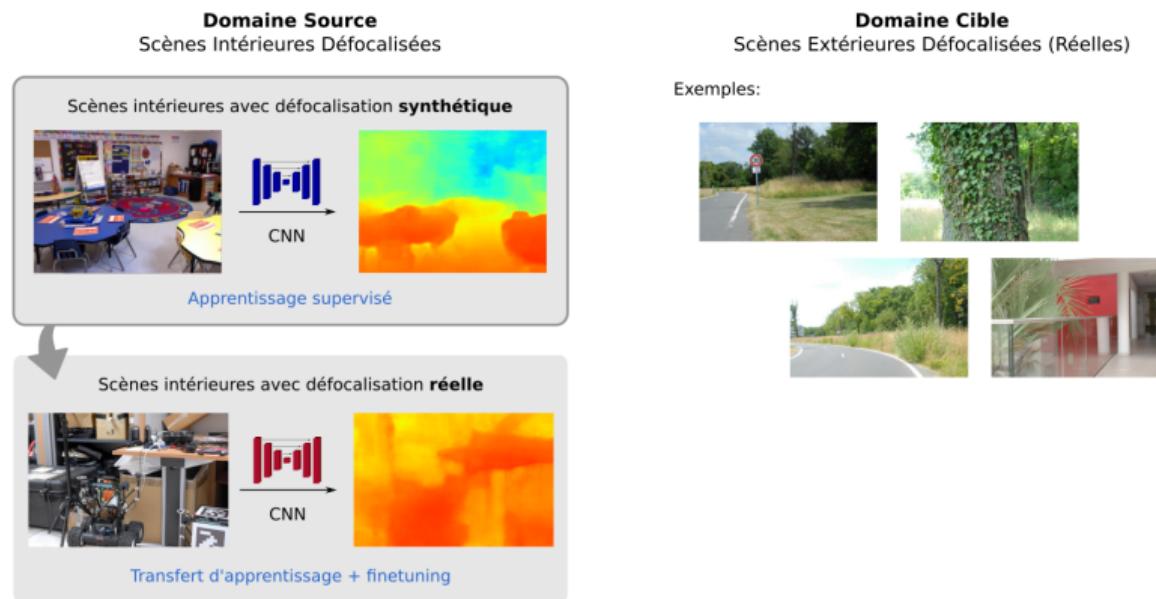
Deep DFD 'in the wild'

Objectif : vérifier la capacité de généralisation du modèle CNN avec des scènes extérieures.



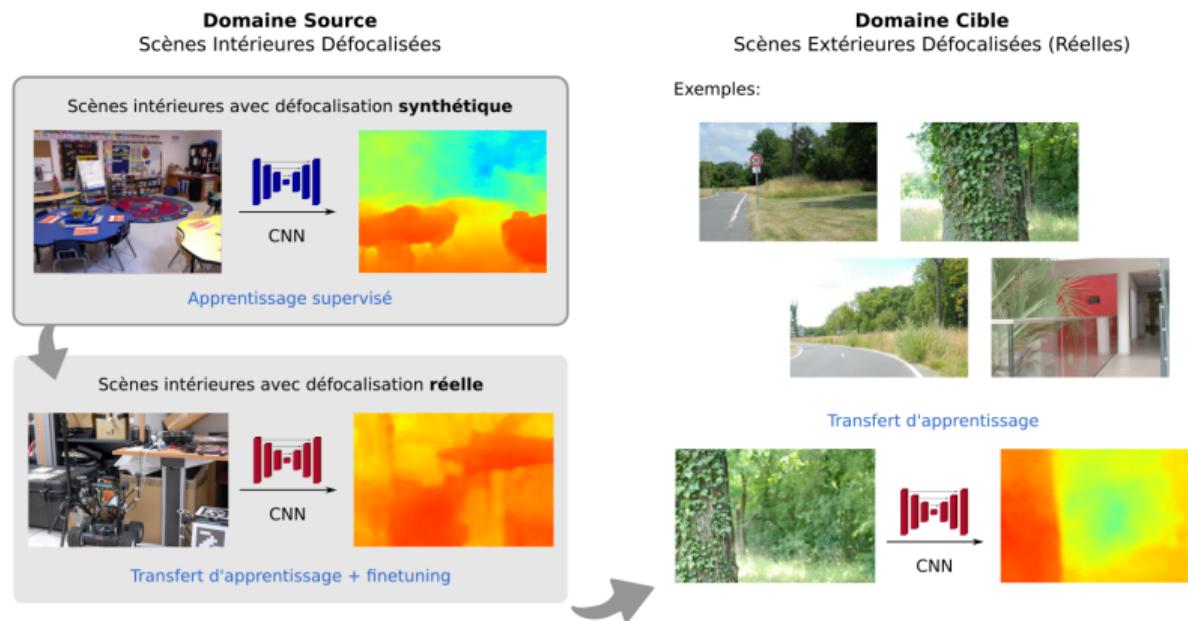
Deep DFD 'in the wild'

Objectif : vérifier la capacité de généralisation du modèle CNN avec des scènes extérieures.

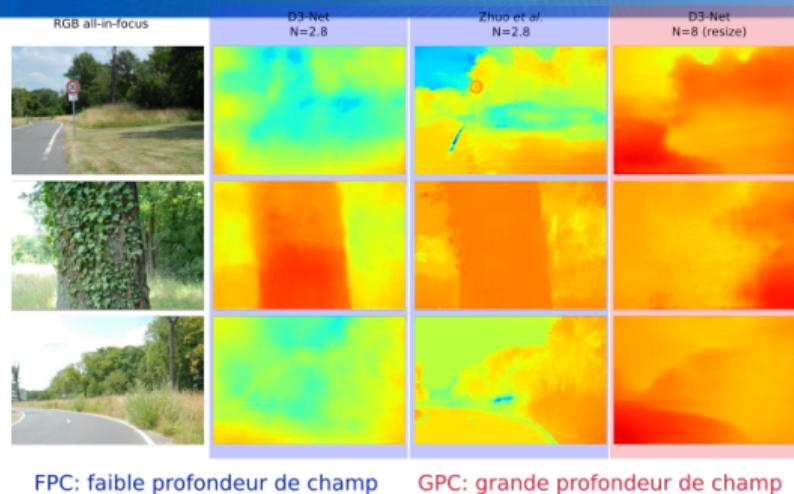


Deep DFD 'in the wild'

Objectif : vérifier la capacité de généralisation du modèle CNN avec des scènes extérieures.



Résultats Deep DFD 'in the wild'



L'apport du flou pour la généralisation de l'apprentissage

- Le modèle Deep-DFD utilise le flou et les aspects géométriques ;
- Zhuo et .al présente des erreurs d'ambiguité ;
- Le modèle entraîné sans flou n'est pas capable de donner des estimations fiables.

→ Carvalho et al., *Deep Depth from Defocus*, ECCV / 3D Rec. in the Wild 2018

Conclusions

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 adverse 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 de grandes bases de données floutées réelles.

Perspectives

- Création d'une base de données avec un capteur DSLR, Intel RealSense, stereo.
- Co-conception d'un capteur avec paramètres optimales pour l'estimation de profondeur avec les CNNs ;



Questions ?

Apport du flou de défocalisation pour
l'estimation de profondeur monoculaire par réseaux de neurones



M. Carvalho, B. Le Saux, P. Trouvé-Peloux, F. Champagnat, A. Almansa

Code : https://github.com/marcelampc/d3net_depth_estimation

Contact : marcela.carvalho@onera.fr <http://mcarvalho.ml>

bertrand.le_saux@onera.fr <http://blesaux.github.io>



Bibliographie succincte

- 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).
- 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., Badrinarayanan, V., & Cipolla, R. (2015). Bayesian segnet : Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv preprint arXiv :1511.02680*.
- 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