

Learning to Discover Cross-Domain Relations with Generative Adversarial Networks

Published by T. Kim, M. Cha, H. Kim, J. K. Lee, and, J. Kim

Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017

Réalisée par:

Ahmed EL-ABEDY

Khalid BOUHOUCH

Encadrée par:

Prof. EL HADRI Aissam

الكلية متعددة التخصصات - ورزازات
+٢٤٣٦٩٤ +٥٢٤٣٦٩٤ - ٥٠٧٠٣٦٧
FACULTÉ POLYDISCIPLINAIRE DE OUARZAZATE



Plan

01

Introduction des réseaux antagonistes génératifs (GAN)

02

Problématique

03

DiscoGAN: Résolution proposée

04

Expériences

05

DiscoGAN: Points forts et Limites

06

DiscoGAN et GANs précédents

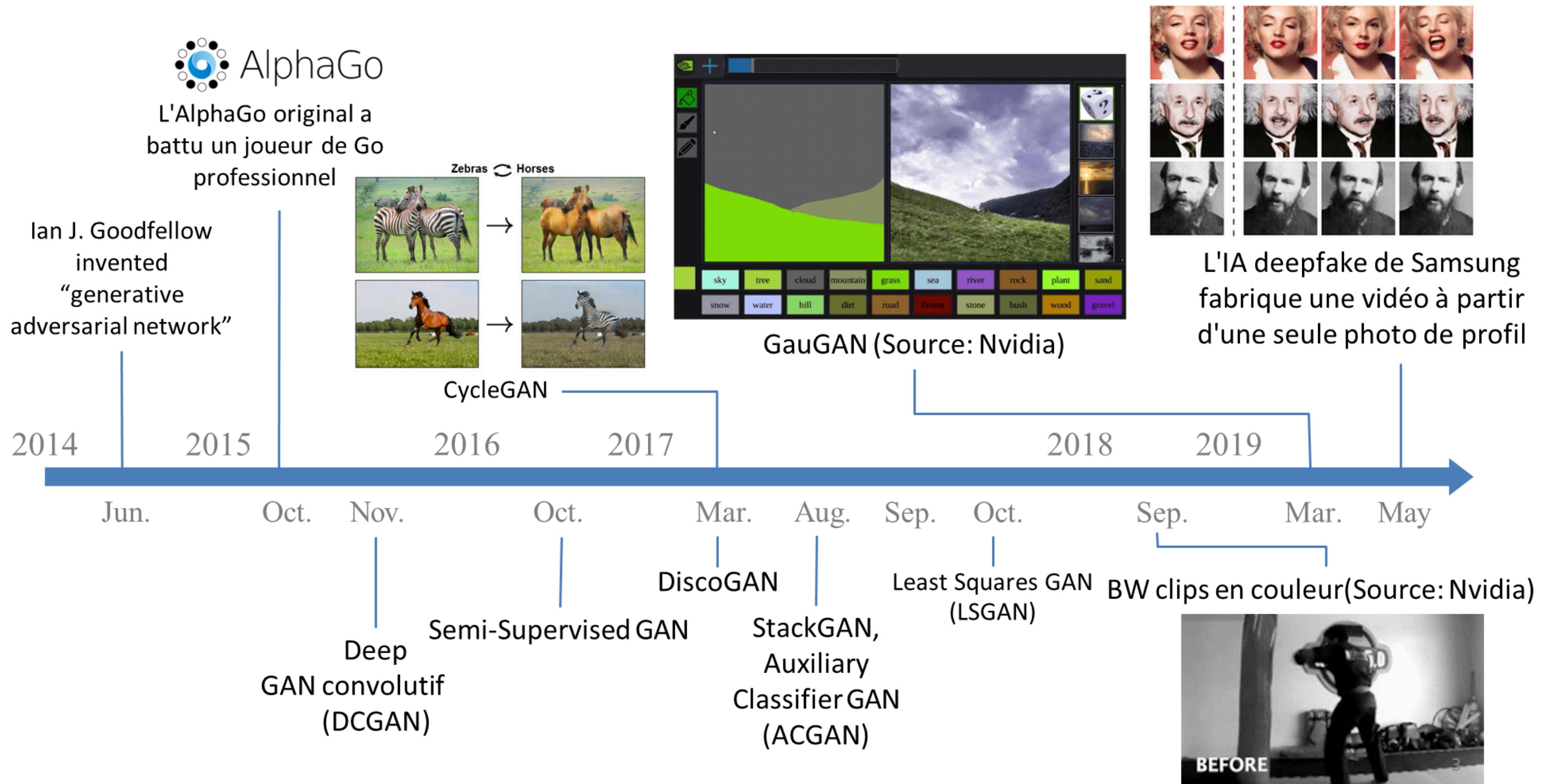
07

Perspectives

08

Conclusion

Introduction: Réseaux antagonistes génératifs (GAN)



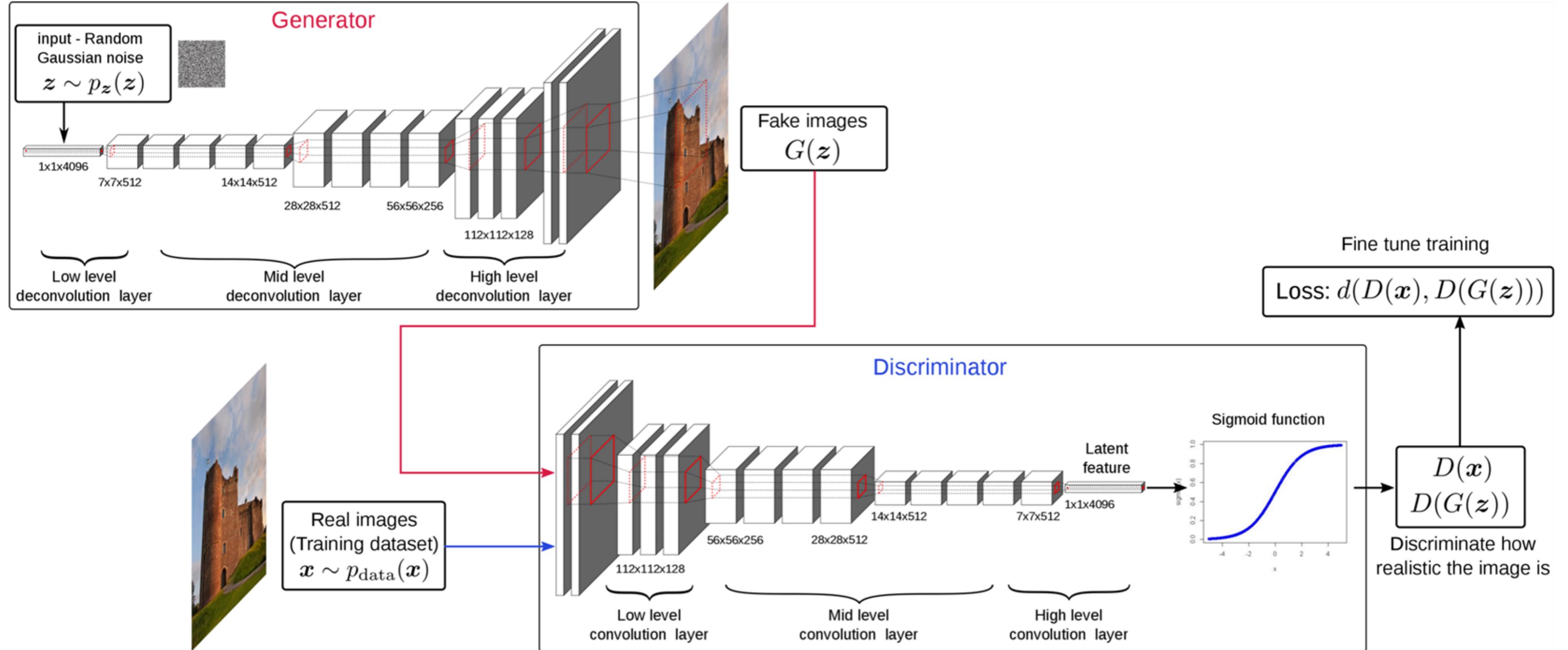
- Avant DiscoGAN, les approches existantes pour relier deux domaines (A et B) avaient des limites
 - Besoin de données appariées
 - Risque de correspondances ambiguës (mode collapse)

Comment traduire d'un domaine à un autre sans exemples directs, tout en s'assurant que la correspondance soit unique et stable ?



DiscoGAN

3-1 Qu'est-ce que DiscoGAN ?

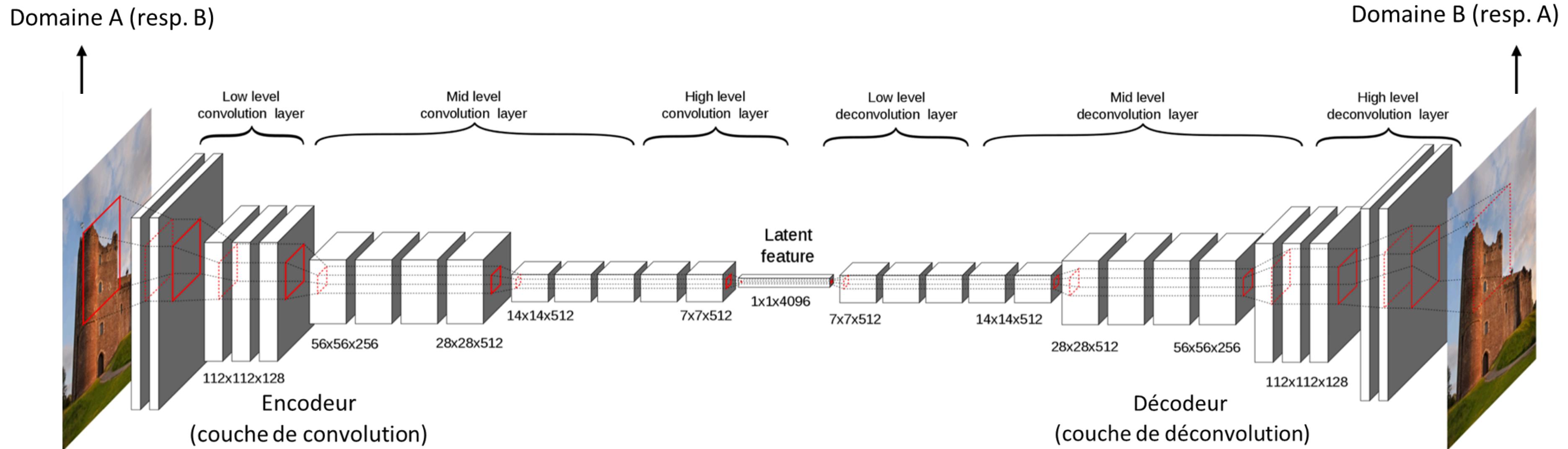


- DiscoGAN est un modèle non supervisé qui apprend à traduire entre deux domaines différents sans données appariées, en assurant une correspondance bijective.

3-2 Architecture DiscoGAN

- **Architecture du générateur**

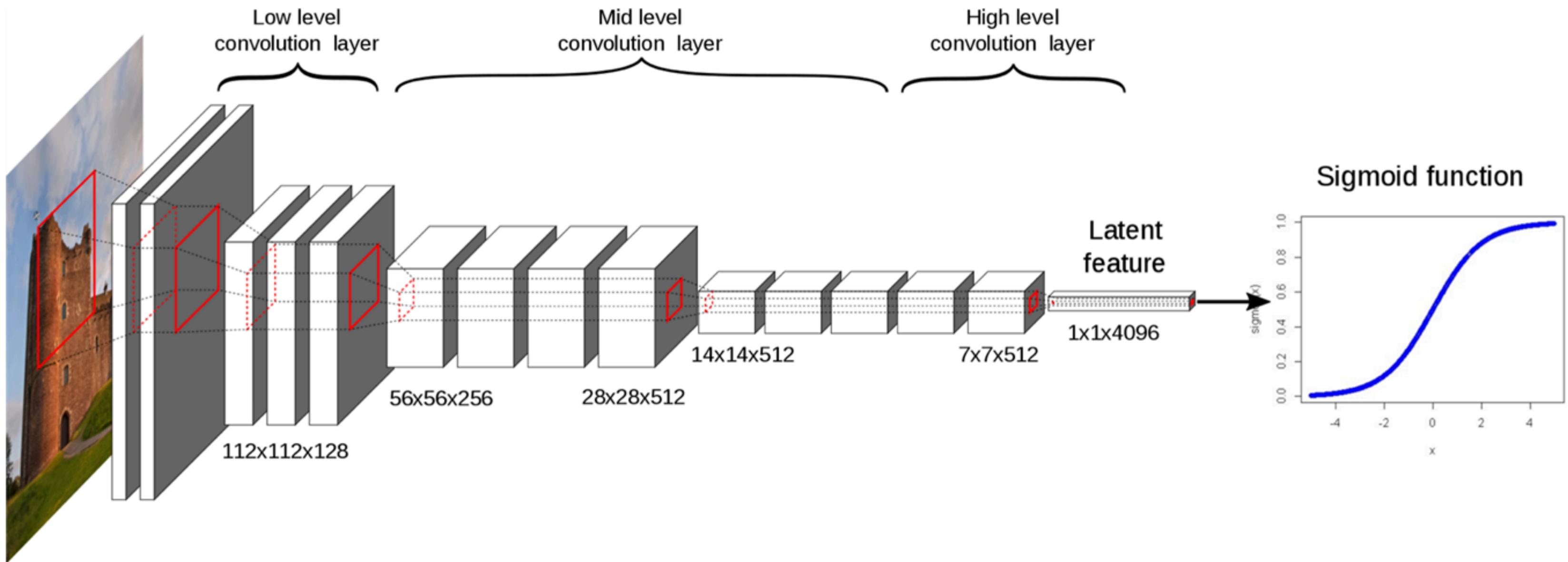
- DiscoGAN repose sur deux générateurs (GAB, GBA) et deux discriminateurs (DA, DB).



Chaque générateur prend une image et la transmet à une paire encodeur-décodeur

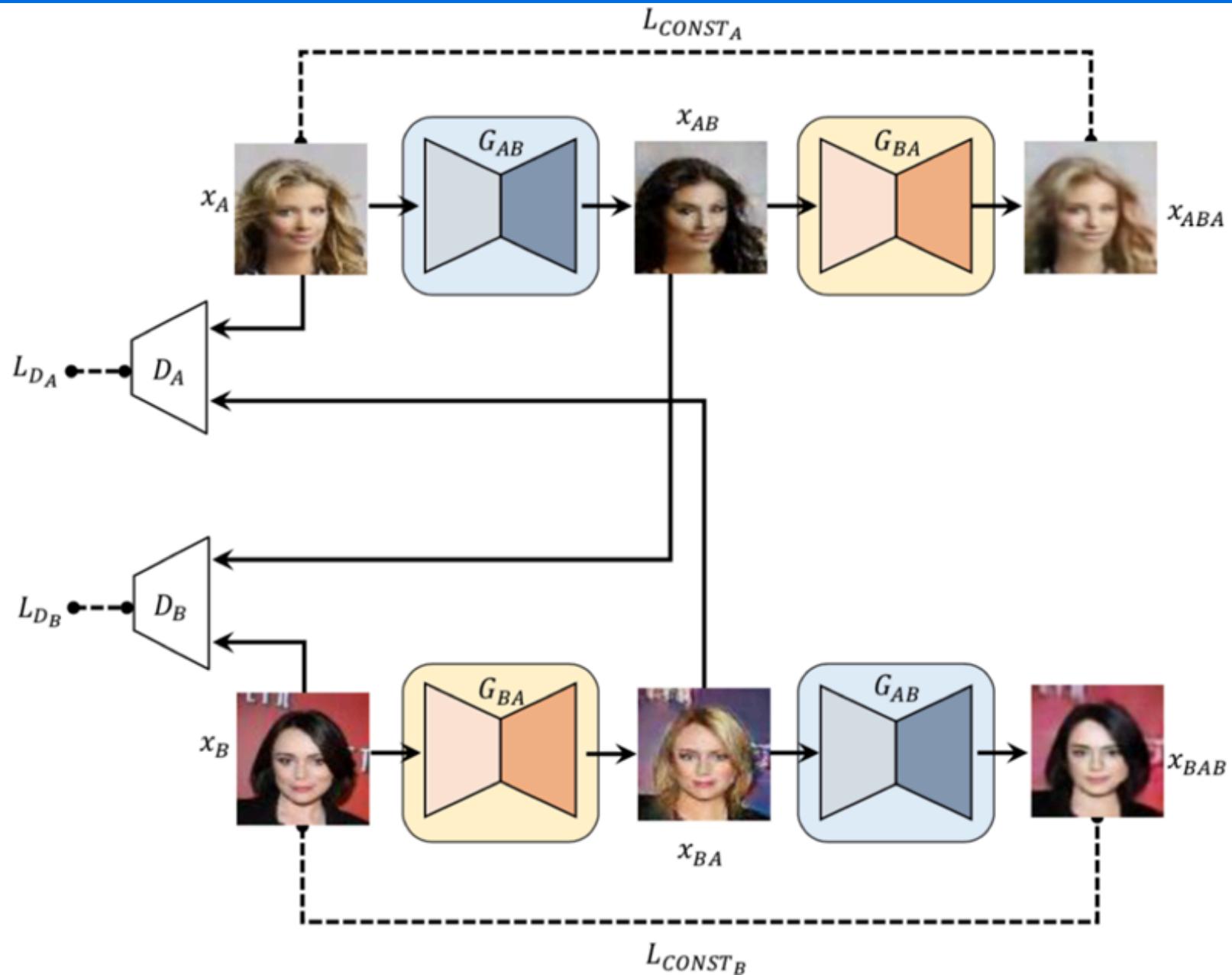
3-2 Architecture DiscoGAN

- Architecture du discriminateur



- Chaque discriminateur alimente une image à travers des couches de convolution.
- Le discriminateur génère une sortie scalaire basée sur la sigmoïde, indiquant le niveau de réalisme de l'image alimentée.

3-3 Formulation du problème



- **Bijection:** ideally $G_{AB}^{>1} = G_{BA}$
 $\rightarrow \min_{G_{AB}}(L_{CONST_A}), \min_{G_{BA}}(L_{CONST_B})$
- **Domain transition:** ideally $x_{AB} \in \mathbb{R}^B, x_{BA} \in \mathbb{R}^A$
 $\rightarrow \min_{D_B}(L_{D_B}), \min_{D_A}(L_{D_A})$

- **La perte de reconstruction** mesure la qualité de la reconstruction de l'entrée d'origine après une séquence de deux générations: $L_{CONST_A} = d(x_{ABA}, x_A)$ comme L_1, L_2 , ou perte de Huber.
- **GAN loss** mesure le réalisme de l'image générée dans le domaine B : $L_{GAN_B} = -\mathbb{E}_{x_A \sim p_A(x)} [\log D_B(x_{AB})]$
- Des contraintes relâchées sont considérées pour garantir la bijection et la transition de domaine

3-3 Formulation du problème

- **Perte adversariale (GAN Loss)**

Chaque générateur essaie de tromper son discriminateur associé :

$$\mathcal{L}_{GAN}^A = \mathbb{E}_{x_A \sim p_A} [\log D_A(x_A)] + \mathbb{E}_{x_B \sim p_B} [\log(1 - D_A(G_{BA}(x_B)))]$$

$$\mathcal{L}_{GAN}^B = \mathbb{E}_{x_B \sim p_B} [\log D_B(x_B)] + \mathbb{E}_{x_A \sim p_A} [\log(1 - D_B(G_{AB}(x_A)))]$$

- **Perte de reconstruction (Cycle-consistency)**

On impose que $A \rightarrow B \rightarrow A \approx A$ et $B \rightarrow A \rightarrow B \approx B$:

$$\mathcal{L}_{recon}^A = \mathbb{E}_{x_A \sim p_A} [d(G_{BA}(G_{AB}(x_A)), x_A)]$$

$$\mathcal{L}_{recon}^B = \mathbb{E}_{x_B \sim p_B} [d(G_{AB}(G_{BA}(x_B)), x_B)]$$

- **Perte totale**

$$\mathcal{L}_{total} = \mathcal{L}_{GAN}^A + \mathcal{L}_{GAN}^B + \lambda(\mathcal{L}_{recon}^A + \mathcal{L}_{recon}^B) \quad \{ \text{où } \lambda \text{ équilibre réalisme et fidélité} \}$$

- **Problème d'optimisation**

$$\min_{G_{AB}, G_{BA}} \max_{D_A, D_B} \mathcal{L}_{total}$$

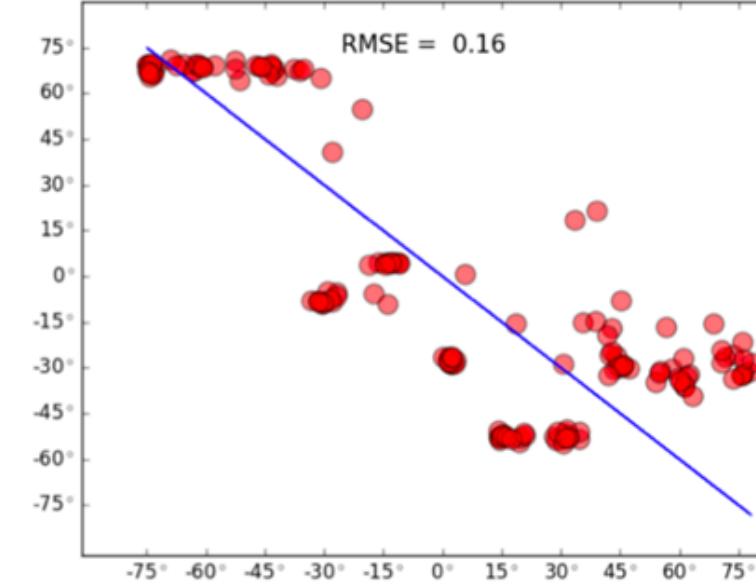
4-1 Voiture à voiture, face à face

Car to car

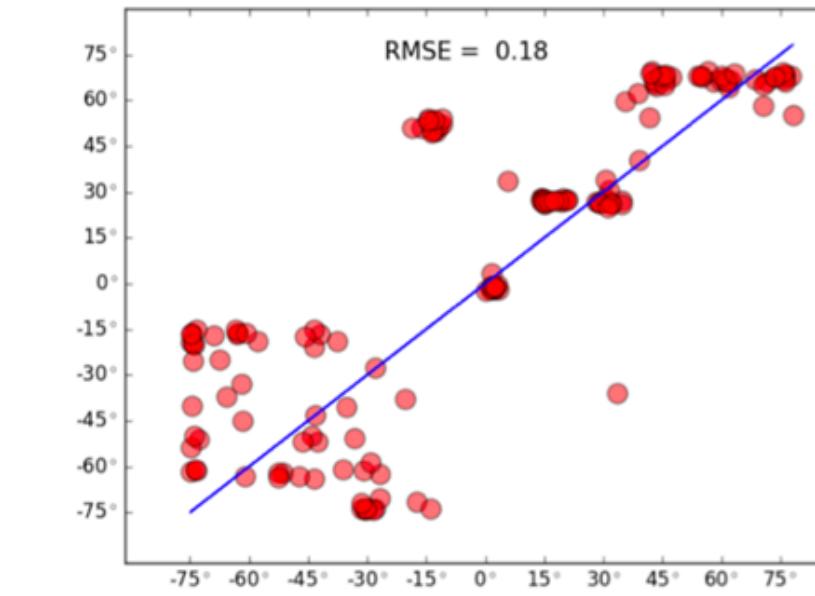
Input data



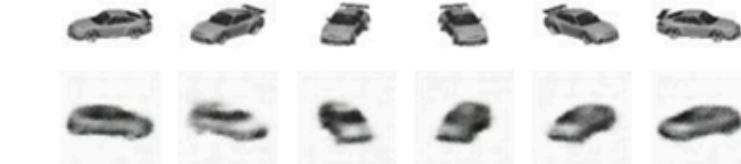
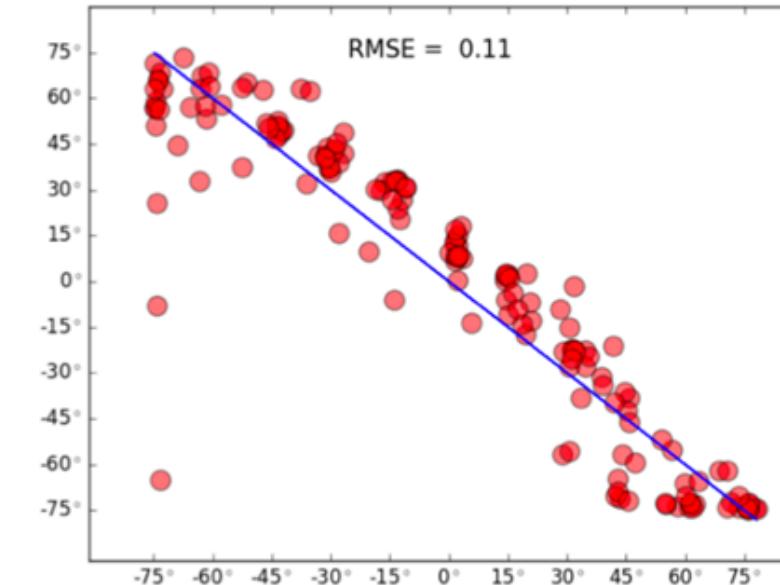
Standard GAN



Face to face

GAN with L_{CONST} 

DiscoGAN



- Tests de reconstruction
- Les résultats de DiscoGAN montrent des corrélations plus élevées (résistance à l'effondrement des modes).

4-2 Domaine réel: Conversion de visages

- Traduction du genre



- Cheveux blonds à noirs,
noirs à blonds



- Des lunettes aux non-lunettes,
des non-lunettes aux lunettes



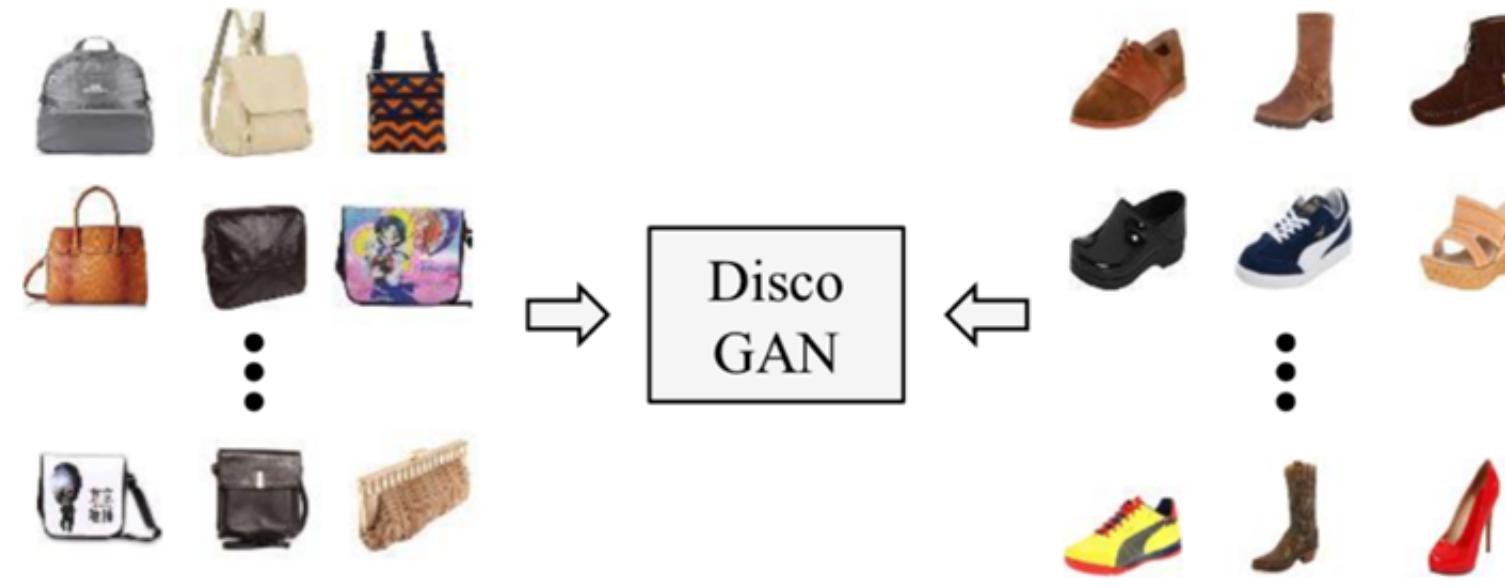
- Traduction d'images de visages masculins en visages féminins et inversement.
- DiscoGAN traduit des traits spécifiques, préservant les autres traits du visage

04

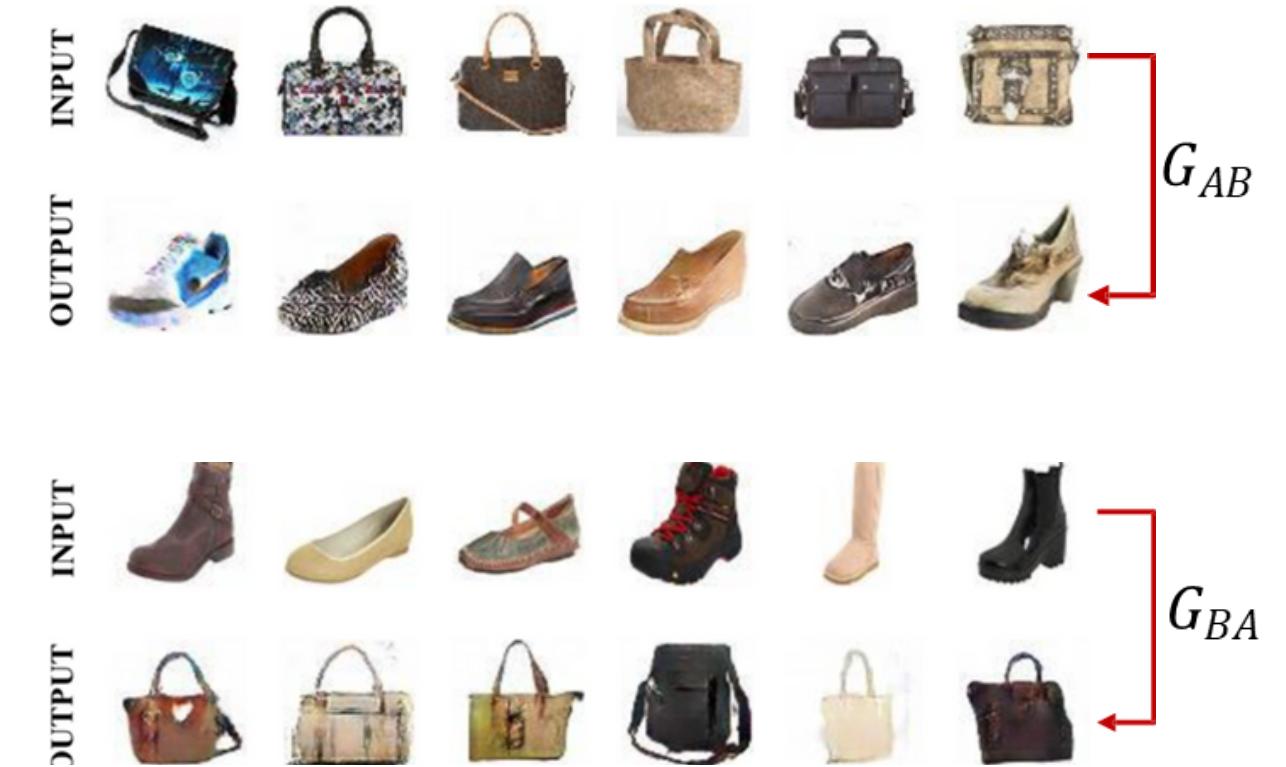
Expérience

4-3 inter-domaines (Sacs à main ↔ Chaussures)

Formation de 2 ensembles de données différents sans étiquetage explicitement apparié



Résultats du transfert de domaine



- DiscoGAN nécessite l'apprentissage de deux ensembles de données différents sans aucune donnée appariée, et ses résultats montrent de meilleures performances et une meilleure résistance au problème d'effondrement de mode.

5-1 Points forts

Non supervision :

- DiscoGAN n'a pas besoin des paires exactes entre les domaines pour apprendre la transformation.
- Il peut donc exploiter des datasets non alignés, ce qui simplifie la collecte des données.

Polyvalence :

- Le modèle fonctionne sur une grande variété de tâches : visages, objets, contours → photos, correspondances abstraites.
- Il n'est pas limité à un type particulier de domaine.

Préservation d'attributs :

- Certaines caractéristiques importantes sont conservées après la transformation, comme l'identité d'un visage ou le style d'un objet.
- Cela garantit que les résultats restent cohérents et crédibles pour l'utilisateur.



5-2 Limites

Résolution faible :

- Les images créées par DiscoGAN sont souvent petites, par exemple 64 ou 128 pixels. À cause de cette faible taille, les petits détails disparaissent ou deviennent flous.

Instabilité GAN :

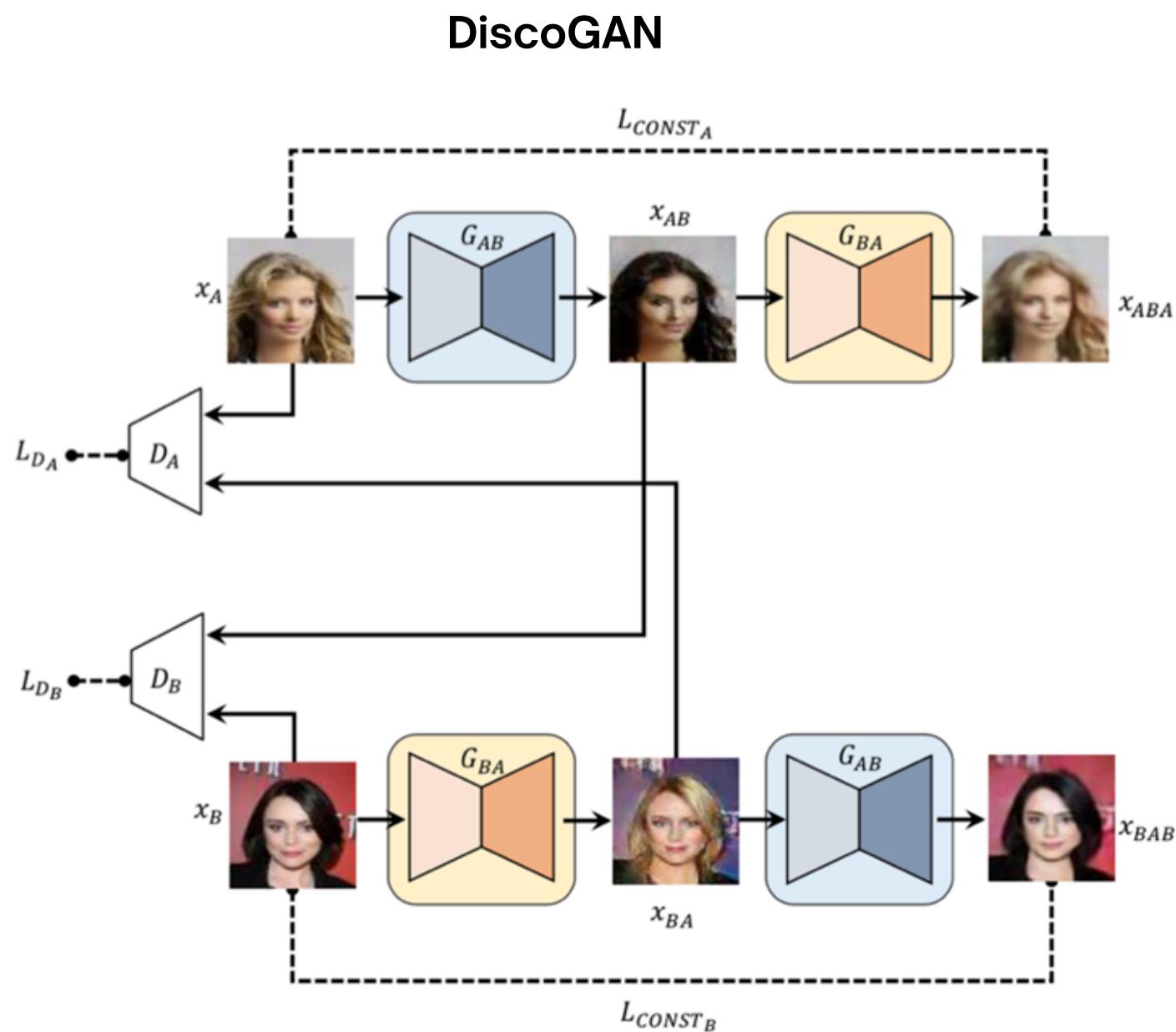
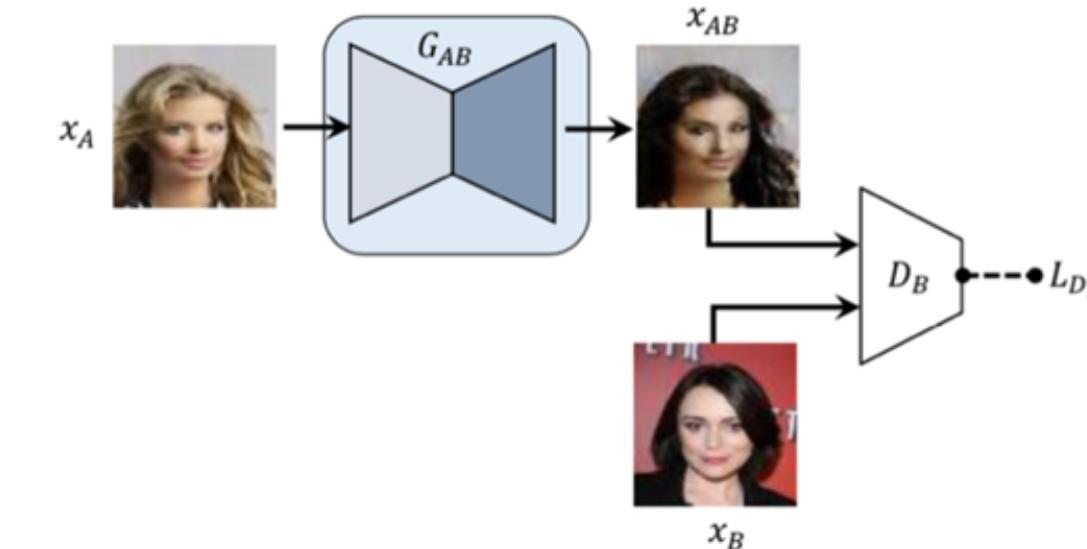
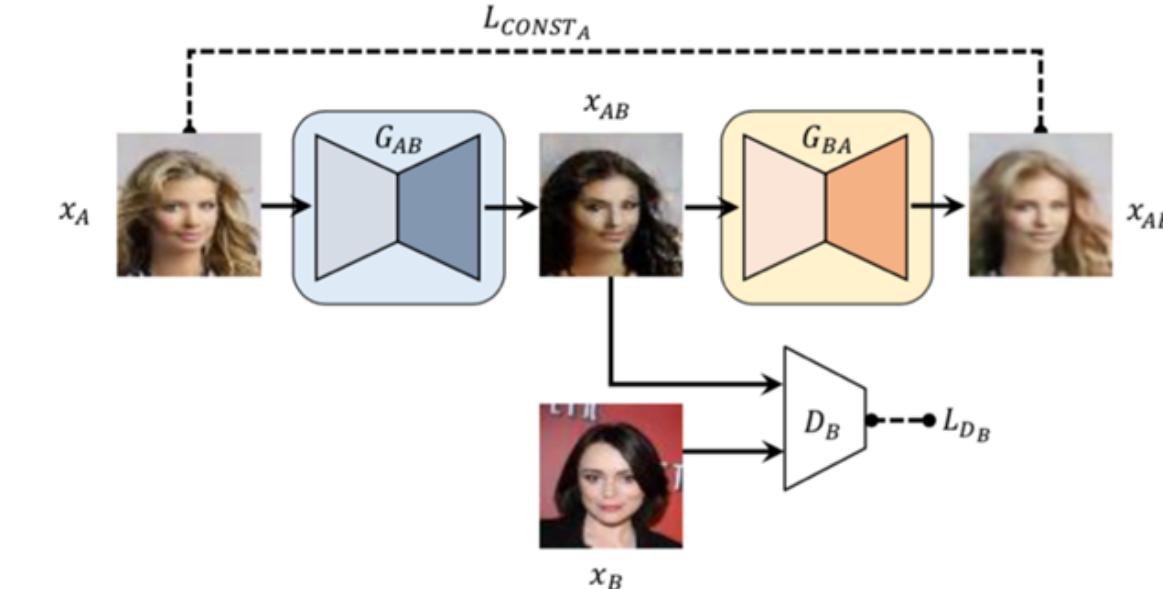
- Comme tous les GANs, DiscoGAN peut être instable lors de l'entraînement.
- Risque de mode collapse, oscillations ou artefacts visuels si les hyperparamètres ne sont pas bien choisis.

Pas multimodal :

- Le modèle génère généralement une seule sortie par entrée, sans diversité dans la transformation possible.
- Il ne capture pas plusieurs styles ou variantes possibles pour un même input.



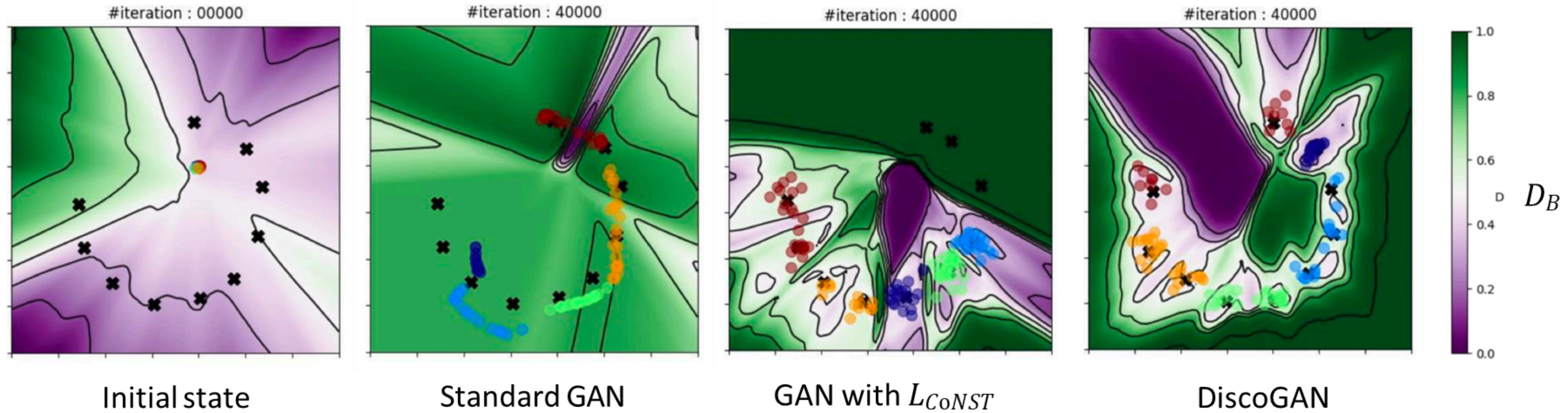
6-1 Comparaison

**Standard GAN with GAN loss****GAN avec une perte de reconstruction et une perte GAN**

- Chaque générateur est constitué d'une paire encodeur-décodeur (l'entrée et la sortie sont des images).
- La perte de GAN (et la reconstruction) doit être minimisée lors des processus d'apprentissage.
- Dans DiscoGAN, deux GAN couplés mappent chaque domaine à son domaine homologue (bijectif).

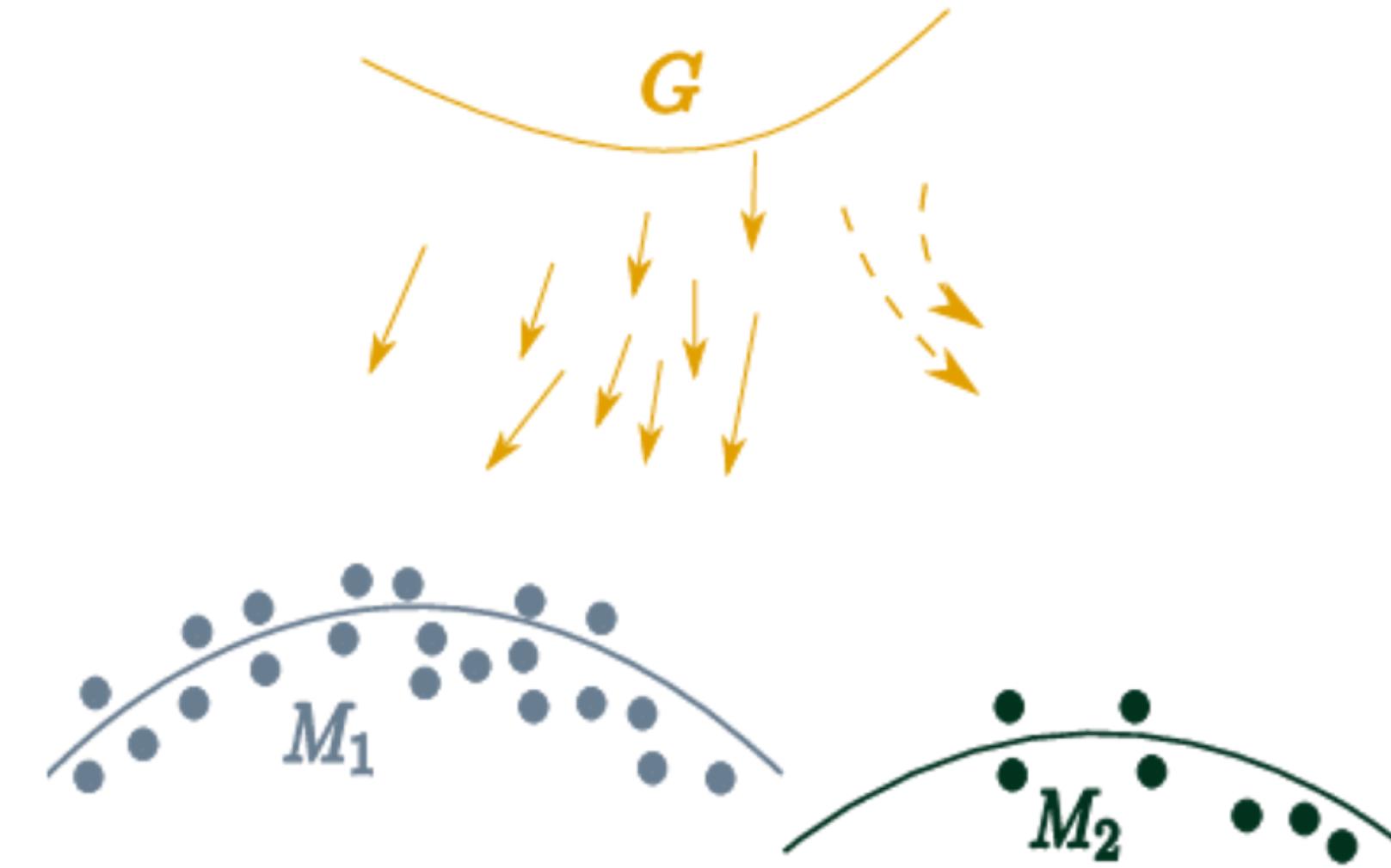
6-2 Problème Collapse

Point colorés:échantillons du domaine A
X noirs : modes cibles du domaine B



- Dans DiscoGAN, le discriminateur B est efficacement trompé par les échantillons traduits du domaine A. Le modèle évite ainsi l'effondrement de mode en générant des traductions réparties dans des régions distinctes et bien délimitées, qui ne se chevauchent pas

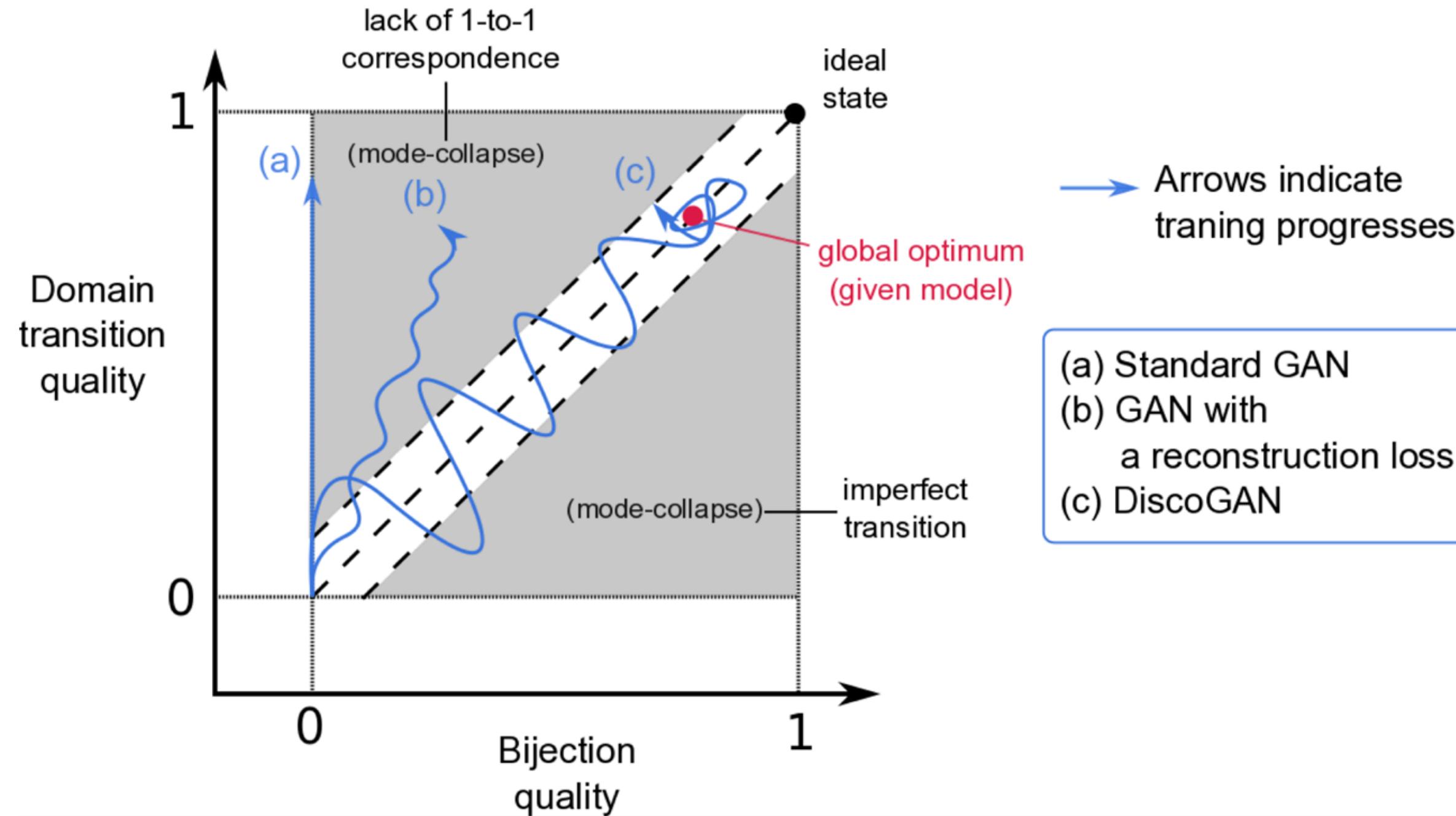
6-2 Problème Collapse



Les gradients sont biaisés vers le mode à partir duquel un plus grand nombre d'échantillons sont tirés pour former les données d'entraînement réelles

- Le générateur génère des images non désirées dans différents modes, ce qui est fréquent dans les GAN.
- Les GAN corrigent généralement ce problème avec des pertes, mais il n'a pas été résolu parfaitement.

6-3 Pourquoi DiscoGAN est-il robuste au mode-collapse ?



- In DiscoGAN, deux modèles couplés sont. Paramètres de partage de G_{AB} et G_{BA}
- Les contraintes des pertes de reconstruction couplées conduisent a la bijection stricte

Vers haute résolution (StyleGAN) :

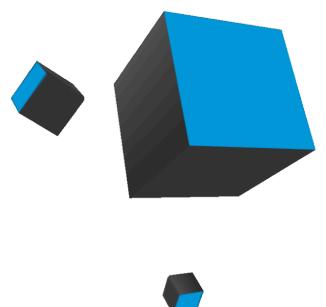
- Intégrer des architectures comme StyleGAN permettrait de produire des images haute résolution et très détaillées.

Vers multimodal (Texte ↔ Image) :

- Ajouter la capacité à générer plusieurs variantes ou traduire depuis d'autres modalités (ex. texte → image) améliorerait la flexibilité.

Vers architectures robustes (CycleGAN, CoGAN) :

- Utiliser des architectures plus stables ou combinées peut réduire les artefacts et améliorer la qualité et la diversité des transformations.
- CycleGAN et CoGAN offrent des mécanismes complémentaires pour la correspondance non supervisée et la cohérence entre domaines.



- DiscoGAN est proposé comme méthode d'apprentissage pour découvrir des relations inter-domaines sans étiquettes de paires.
- Les résultats ont montré de meilleures performances et une meilleure robustesse à l'effondrement de modes. La symétrie obtenue par le couplage de deux GAN est considérée comme un facteur clé de la robustesse dynamique



Références

- Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to dis-cover cross-domain relations with generative adversarial networks. 5 2017.-
- Le point crucial de la présentation
T. Kim, et al., Learning to Discover Cross-Domain Relations with Generative Adversarial Networks (arXiv) ... <https://arxiv.org/abs/1703.05192>
-Recent generative technologies
Apple annonce Animoji (The Verge) ... <https://www.theverge.com/2017/9/12/16290210/new-iphone-emoji-animated-animoji-apple-ios-11-update>
L'IA peut convertir des clips en noir et blanc en couleur(Nvidia Developer) ... <https://news.developer.nvidia.com/ai-can-convert-black-and-white-clips-into-color/>
<https://www.theverge.com/2019/5/23/18637373/deepfakes-samsung-ai-research-results-single-photo-algorithm>
c (The Verge) ... <https://www.theverge.com/2019/6/20/18692671/deepfake-technology-singing-talking-video-portrait-from-a-single-image-imperial-college-samsung>
-Les législateurs américains affirment que les deepfakes de l'IA « ont le potentiel de perturber toutes les facettes de notre société » (The Verge) ...
<https://www.theverge.com/2018/9/14/17859188/ai-deepfakes-national-security-threat-lawmakers-letter-intelligence-community>
D. Vint, M. Anderson, Y. Yang, C. Ilioudis, G. Di Caterina, and C. Clément. Automatique cible
reconnaissance pour faible résolution images sar pénétrant le feuillage à l'aide de cnn et de gans.
Télé-détection, 13 :596, 2021
Christian Janiesch, Patrick Zschech, and Kai Heinrich. Machine learning and deep learning.
Electronic Markets, 31 :685–695, 9 2021
aurence Moroney. IA et apprentissage automatique pour les codeurs. O'Reilly Media, Inc.,
October 2020.

Thank you

