

Génération de sprites Pokémon avec VAE et GAN

Wassim Chikhi

Master 2 Vision et Machine Intelligente – 2025/2026

Notebook rendu : `GANLab_To_complete2.ipynb`

1. Introduction

L'objectif de ce TP est d'étudier expérimentalement deux approches de génération d'images appliquées à des sprites Pokémon : les **Variational Autoencoders (VAE)** et les **Generative Adversarial Networks (GAN)**.

Le VAE vise à apprendre un espace latent continu permettant la reconstruction et l'échantillonnage d'images, tandis que le GAN repose sur un apprentissage adversarial entre un générateur et un discriminateur afin de produire des images visuellement réalistes.

L'accent est mis sur l'analyse expérimentale : influence des hyperparamètres, stabilité de l'apprentissage et qualité visuelle des images générées.

2. Données et environnement

Le dataset est constitué de sprites Pokémon RGB, redimensionnés en 64×64 et normalisés. Toutes les expériences ont été réalisées dans un notebook Jupyter en Python, en utilisant PyTorch et une accélération GPU lorsque disponible.

3. Meilleurs résultats (présentés en premier)

3.1. Meilleur modèle GAN (DCGAN)

Paramètres

Le meilleur compromis entre stabilité et qualité visuelle a été obtenu avec :

- architecture : DCGAN ;
- dimension du bruit : **latent_dim = 100** ;
- optimiseur : Adam ($\beta_1 = 0.5$, $\beta_2 = 0.999$) ;
- learning rate : 2×10^{-4} ;
- nombre d'époques : 20 à 30.

Courbes de loss et images générées

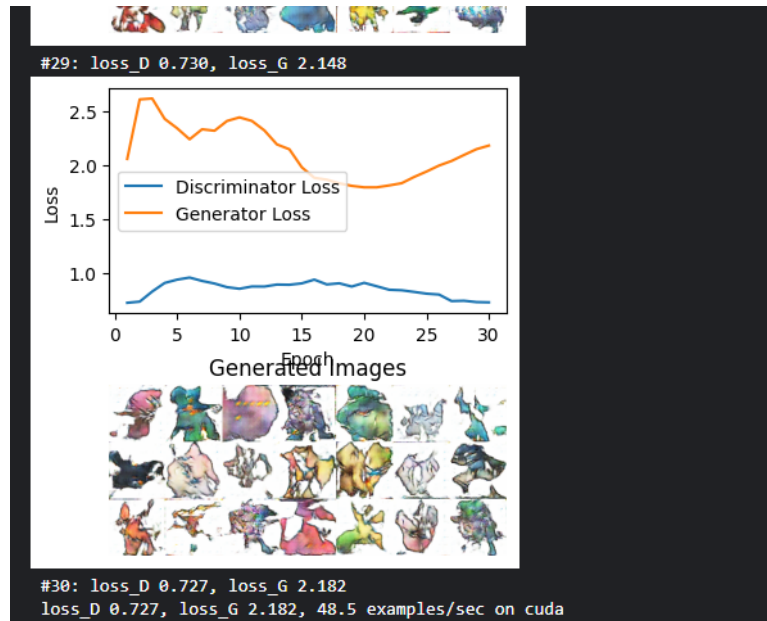


FIGURE 1. Meilleur résultat GAN : courbes de loss du générateur et du discriminateur, accompagnées d'une grille d'images générées (epoch ≈ 30).

Les courbes montrent un entraînement stable : la loss du discriminateur reste proche de 1, tandis que la loss du générateur se stabilise autour de 1.6–2.2. Cela indique que le discriminateur n'écrase pas le générateur, permettant à ce dernier d'apprendre efficacement.

Analyse visuelle

Les images générées présentent des silhouettes cohérentes et des palettes de couleurs plausibles rappelant clairement des sprites Pokémon. Les détails fins restent imparfaits, mais le rendu est nettement supérieur à celui obtenu avec le VAE.

4. Étude des paramètres et stabilité du GAN

4.1. Cas instable : learning rate trop élevé

Avec un learning rate plus élevé (par exemple 10^{-3}), le discriminateur apprend trop rapidement. La loss du discriminateur chute alors vers 0 tandis que la loss du générateur augmente fortement, ce qui conduit à des images générées essentiellement bruitées.

4.2. Runs stabilisés

En réduisant le learning rate à 2×10^{-4} , l'apprentissage devient plus équilibré. Les pertes oscillent dans une plage raisonnable et la qualité visuelle des images s'améliore progressivement.

4.3. Choix de l'époque retenue

Sur les runs plus longs, on observe parfois une dégradation après un certain nombre d'époques, lorsque le discriminateur reprend l'avantage. Les meilleurs résultats visuels correspondent donc à une époque intermédiaire, justifiant un arrêt anticipé.

5. Résultats VAE

5.1. Architecture et apprentissage

Le VAE utilisé est convolutionnel, avec un espace latent de dimension 128. La fonction de coût combine une erreur de reconstruction (MSE) et une divergence de Kullback–Leibler, assurant la régularisation de l'espace latent.

5.2. Reconstructions

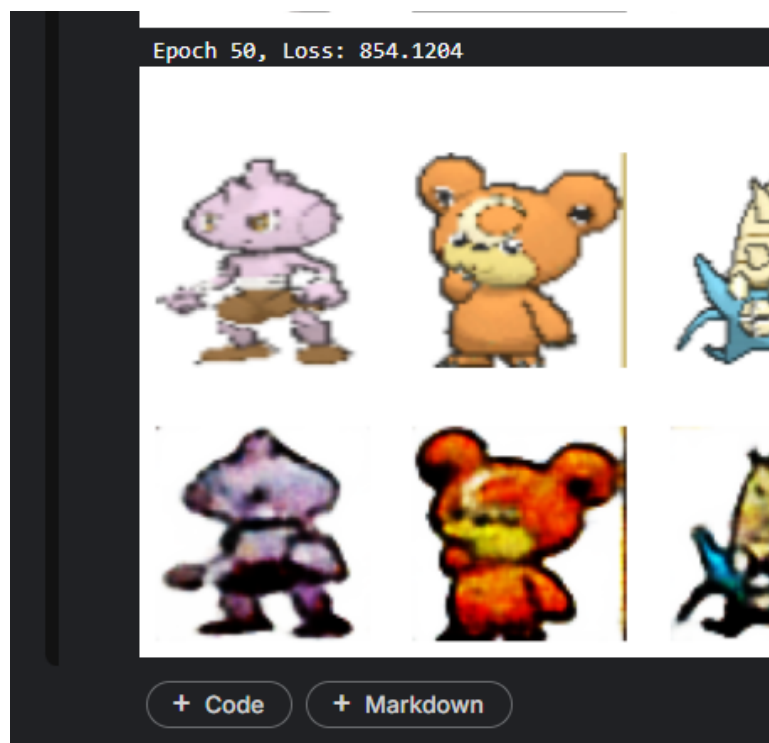


FIGURE 2. Reconstructions VAE à l'époque 50 : images originales (haut) et reconstructions correspondantes (bas).

Les reconstructions reproduisent correctement la structure globale et les couleurs principales, mais restent floues, ce qui est une limitation classique des VAE.

5.3. Échantillonnage

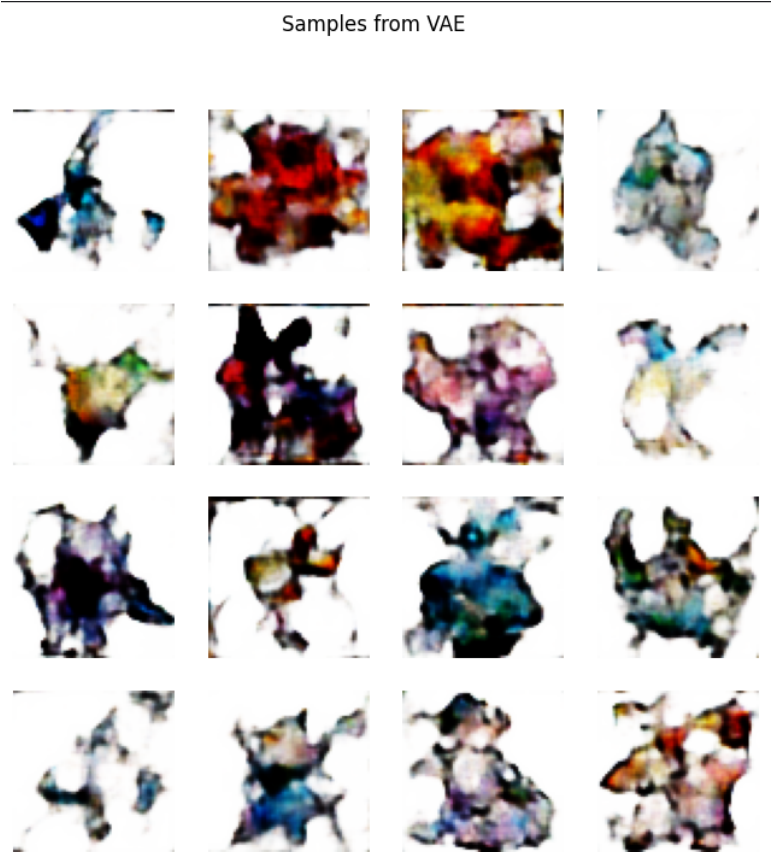


FIGURE 3. Échantillons générés par le VAE à partir de vecteurs latents tirés selon $\mathcal{N}(0, I)$.

Les images générées présentent des amas colorés avec parfois une structure reconnaissable, mais un réalisme inférieur à celui du GAN.

6. Essais VAE préliminaires

Avant de fixer les paramètres finaux, plusieurs essais rapides ont été réalisés afin d’explorer l’influence du learning rate, de la dimension latente et du nombre d’époques. Ces essais n’ont pas tous été sauvegardés visuellement, mais les valeurs de loss ont permis d’orienter les choix finaux.

| Learning rate | latent_dim | Époques | Loss finale |
|--------------------|------------|---------|-------------|
| 1×10^{-3} | 248 | 10 | 1010.21 |
| 0.002 | 248 | 20 | 933.10 |
| 0.001 | 128 | 25 | 892.75 |

TABLE 1. Essais VAE préliminaires utilisés pour guider le choix des hyperparamètres.

7. Comparaison VAE / GAN

| Critère | VAE | GAN |
|------------------|---------------|-------------------------|
| Stabilité | Bonne | Sensible aux paramètres |
| Netteté | Faible (flou) | Meilleure |
| Qualité visuelle | Limitée | Supérieure |

TABLE 2. Comparaison qualitative entre VAE et GAN.

8. Conclusion

Ce TP met en évidence l'importance du choix des hyperparamètres dans les modèles génératifs. Le VAE offre une grande stabilité et de bonnes reconstructions, mais génère des images floues en génération libre. Le GAN, bien que plus délicat à entraîner, produit des images visuellement plus réalistes lorsque l'apprentissage est correctement équilibré.

Annexe – Liste des figures

| Figure | Fichier | Description |
|--------|-----------------------|--------------------------------------------------------------|
| 1 | gan_best_epoch30.png | Courbes de loss du GAN et images générées (meilleur run). |
| 2 | vae_recon_epoch50.png | Reconstructions VAE : originaux vs reconstructions. |
| 3 | vae_samples.png | Échantillons générés par le VAE à partir de l'espace latent. |