Tests & Comparaison : Architecture I2T vs T2T

Évaluation des modèles MV-Adapter

Par Kevin Voisin et Marco Caporizzi

1. Introduction

Lors de la revue de la semaine 4 du projet Duckify, Nous avons présenté au reste de l'équipe notre évaluation & conclusion des différentes approches pour générer des contours et des textures sur un modèle 3D de canard, avec pour objectif final un traçage physique par un robot. Ce rapport détaille l'évaluation technique de deux architectures de génération basées sur MV-Adapter : l'approche Image-to-Texture (I2T) et l'approche Text-to-Texture (T2T).

2. Approche 1 : Image-to-Texture (I2T)

L'approche Image-to-Texture repose sur un processus divisé en deux phases distinctes.

2.1. Description du Pipeline

- **Phase 1 (Génération d'Image) :** Utilisation du modèle FLUX KONTEXT 1 (32 GB VRAM) pour générer une image 2D à partir d'un prompt textuel et d'une image du maillage (mesh).
- **Phase 2 (Projection en Texture) :** Utilisation d'un modèle de diffusion SDXL (32 GB VRAM) pour transformer l'image 2D générée en Phase 1, combinée au maillage et à la carte UV, en une texture 3D finale.

2.2. Limites et Inconvénients Identifiés

Bien que cette méthode offre des résultats plus déterministes, nos tests ont mis en évidence plusieurs problèmes majeurs :

- **Temps de génération et taux d'échec :** Avec une solution sur site, la génération d'une seule image prend jusqu'à 3 minutes. Le taux de succès de cette première phase étant inférieur à 20 %, il faut compter environ 15 minutes pour obtenir une image acceptable. De plus, le taux de succès global de la Phase 2 s'effondre à 10 %, rendant cette méthode inadaptée.
- **Incohérences Multi-vues :** Utiliser une seule image en entrée force le modèle à deviner les surfaces invisibles. Cela entraîne des incohérences visuelles importantes selon l'angle de vue du rendu 3D.
- **Fuites de fond (Background Leakage) :** Les images générées par FLUX manquent de constance, et les éléments du décor « fuitent » fréquemment sur la surface du modèle. Plus le motif demandé est complexe, plus ce phénomène de fuite est sévère.
- **Faible variance des résultats** : Sachant qu'il se base sur une image déja toute prete, le spectre des choses & textures « nouvelles » est très faible

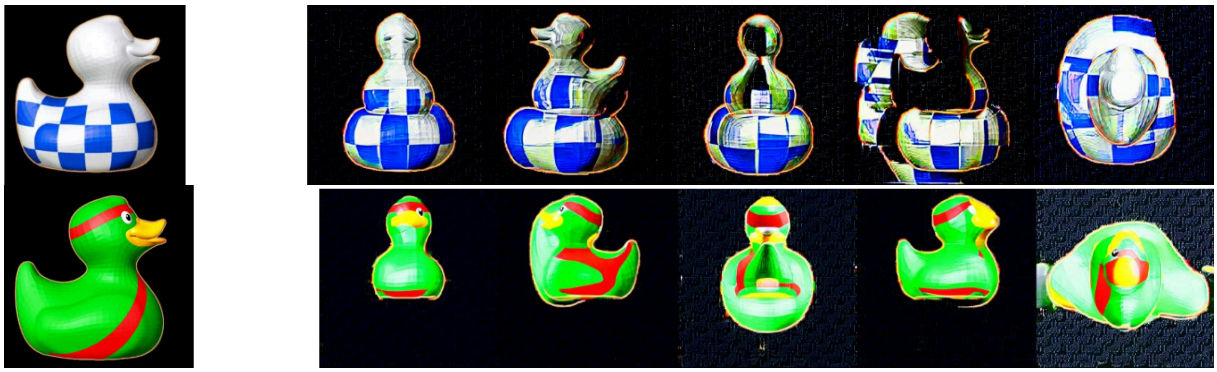


Fig. 1. – MV-Adapter 6 views generated : Background leakage

3. Approche 2 : Text-to-Texture (T2T)

Afin de contourner les limitations strictes de l'I2T, nous avons évalué un pipeline direct Text-to-Texture (T2T) utilisant le modèle de diffusion SDXL.

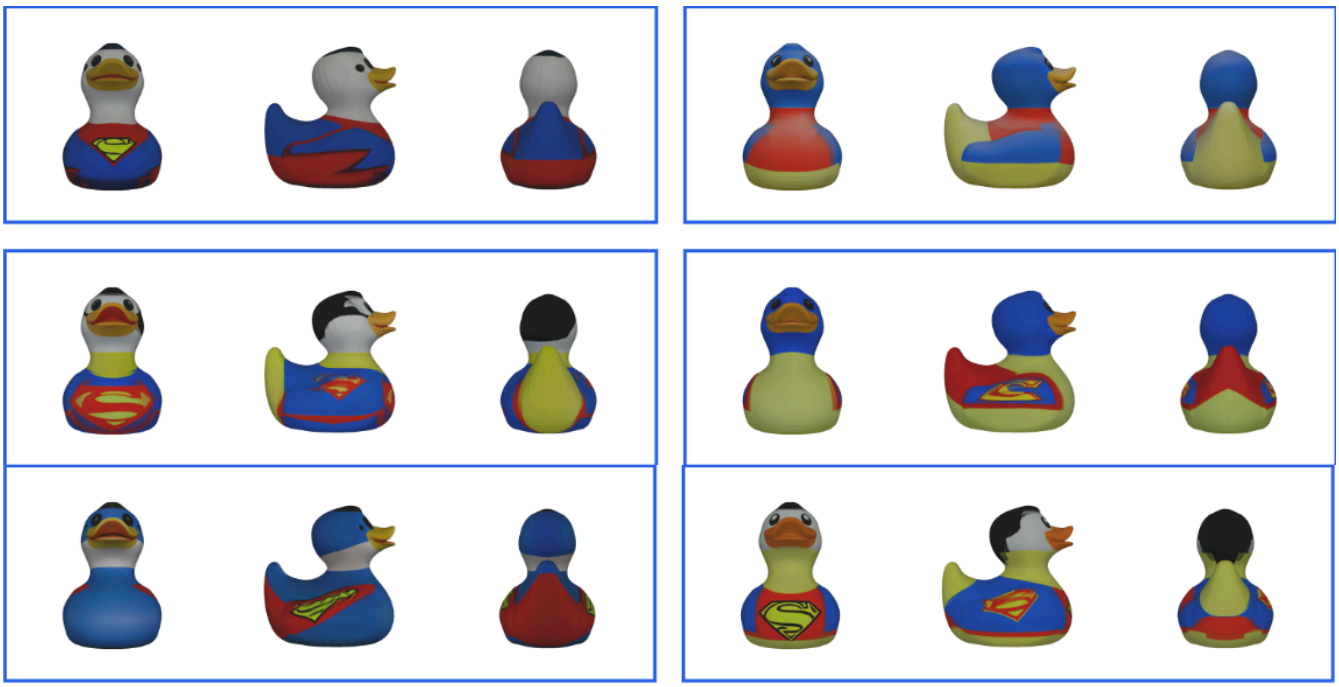
3.1. Performances et Avantages

L'approche T2T simplifie grandement le processus et résout les points bloquants de la méthode précédente :

- **Vitesse accrue :** Le temps de génération est réduit à seulement 45 secondes par texture. Il ne faut désormais plus que 1 minute et 30 secondes en moyenne pour obtenir un bon résultat.
- **Amélioration visuelle :** Le phénomène de « Background Leakage » (fuite de fond) est considérablement réduit par rapport à l'I2T.

3.2. Inconvénient observé

Malgré ces nettes améliorations en termes de temps et de propreté globale, l'approche T2T présente un défaut : elle introduit **davantage de variance dans les résultats** générés. Cette particularité pourra être transformé en force : imaginons faire choisir au client sa génération préférée , ou concevoir un système d'évaluation de la qualité sur la texture et potentiellement choisir la/les meilleure(s) des 5-10 générations



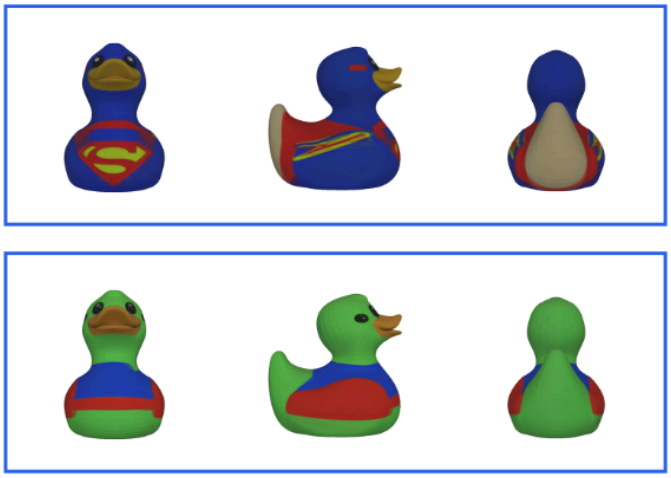


Fig. 2. – T2T - Variance des résultats , Prompt : « A superman duck »

4. Conclusion et Synthèse

L'évaluation des deux pipelines a permis de mettre en évidence des différences drastiques d'efficacité. Le tableau ci-dessous résume les métriques clés de notre benchmark :

Critère	I2T (Image-to-Texture)	T2T (Text-to-Texture)
Fuite du fond (Leakage)	Sévère	Très faible
Inconstance (Variance)	Incohérences multi-vues	Plus de variance dans les résultats
Temps par génération	3 minutes	45 secondes
Temps pour un bon résultat	15 minutes	1m-5mn
Déterminisme	Résultats plus déterministes	-

Choix final : L'architecture **Text-to-Texture (T2T)** a été retenue par l'équipe GenAI. Bien qu'elle induise une plus forte variance dans les résultats générés, ses avantages en termes de vitesse d'exécution (45 secondes contre 3 minutes) et la quasi-disparition des fuites de fond en font la seule solution viable à long terme pour s'intégrer au pipeline global du robot.

5. Sources & Crédits

- Page de projet MVAdapter
- *Utilisation de Gemini afin de formaliser/structurer plus rapidement le contenu de la présentation du CTO de la semaine 4 voir ici*