# 聯大資工系專題製作研討會議之論文

# 進化中的3D 模型：以基因演算法最佳化從影像生成之3D 模型

# Evolutionary 3D Models: Refining 2D-to-3D Model Generation by Genetic Algorithm


**温育瑋 教授**

**李鑑軒、郭崇銘、卓政弘、黃仔彣**

**國立聯合大學 資訊工程學系**

苗栗市南勢里聯大2號

yuwei.james.wen@gmail.com



## 摘要

由於3D 模型的生成在動畫、遊戲開發與虛擬現實等領域需求持續增長，而現有技術在2D 圖像生成上的成熟度與其在3D 生成上的形成鮮明對比。以 Shap-E 為代表的生成模型，儘管能透過單張圖像生成3D 模型，但由於僅能捕捉單一視角，對物體的背面或隱藏部分多依賴模型推測，進而導致生成結果常與輸入圖像存在差異。為解決這一問題，本研究提出結合基因演算法（Genetic Algorithm, GA）的方法，透過多張輸入圖像最佳化從影像生成之3D 模型，提升 Shap-E 對物體全貌的重建能力。GA 以輸入圖像與3D 模型多視角渲染圖的相似度作為適應性函數，最佳化生成結果，避免直接調整模型結構所需的高計算成本。這一方法在提升生成準確性的同時，降低了技術應用的門檻，推動3D 生成技術更廣泛的實用化與發展。


## 一、緒論

本研究聚焦於從2D 圖像生成3D 模型的技術改進，旨在解決現有生成技術中輸出的3D 模型與輸入圖像差異過大的問題。現有的3D 生成式人工智慧技術，雖具一定潛力，但因複雜性與高成本限制了其普及性與應用價值。

本研究以 Shap-E 模型為基礎，充分利用其處理多樣化圖像類型的能力，並結合基因演算法（Genetic Algorithm, GA）進行生成結果的最佳化，目的是提高輸出模型與輸入圖像之間的相似度，進而提升技術的實用性與應用範圍。同時，本研究聚焦於潛空間向量的最佳化，該向量作為資料結構的高維表示，在生成模型的精確性與細節起到關鍵作用。

研究內容包括比較不同生成模型、測試圖像相似度評估方法，以及調整參數進行效能最佳化。通過改進技術，本研究期望提升從2D 圖像生成3D 模型的效率和品質，並推動其在創意產業中的應用，例如動畫製作和遊戲開發。藝術家可快速將2D 角色設計轉換為3D 模型，節省時間與成本；遊戲開發者則能創造多樣化的場景與角色，提高遊戲的沉浸感與參與度。

## 二、文獻探討

近年來，3D 模型的生成技術取得顯著進展。特別是在生成式人工智慧的應用上，如 Shap-E [10] 使用基於 Transformer [1] 的編碼器來處理文本和圖像輸入，將其轉換為高維特徵向量

作為潛空間向量，以潛空間向量當作多層感知器（Multilayer Perceptron, MLP）[9] 的參數，最終生成出3D 模型。Shap-E 的優點包括能夠產生多種形狀和紋理，計算速度較快，並且能選擇輸出的3D 模型為網格或 Neural Radiance Fields（NeRF）[3] 表示形式。NeRF 的優點在於生成的結果在不同視角下較一致，非常適合表示真實存在的場景。然而缺點為需要輸入多張圖像並要求標註角度，且生成的結果不適合直接匯入如 Blender [4] 等3D 建模軟體進行後續的處理。在 Multiview Neural Surface Reconstruction by Disentangling Geometry and Appear ance [16] 中，該研究使用多視角的2D 圖像重建3D 表面，利用 MLP 學習使用 Signed distance function（SDF）[12] 表示幾何形狀，得出幾何形狀後從相機位置向場景中投射光線，確定光線與幾何表面的交點，在訓練過程中從每張圖像中隨機抽樣像素，並從中取得像素的訊息，包括顏色、相機位置和方向，然後最佳化損失函數，以找到幾何形狀和渲染器的參數，該方法在多視角圖像的訓練下表現出良好的廣義化能力，能夠從不同視角下的圖像中學習到一致的3D 形狀，然而其缺點也是相當依賴輸入多張圖像並 要求標註角度。同樣使用潛空間向量來生成結果的有 ProlificDreamer [18] ，這篇研究將隨機初始化的3D 模型渲染為2D 圖像加上雜訊，然後使用 Diffusion Model [13] 將輸入的文字產 生2D 圖像並使用 Low-Rank Adaptation of Large Language Models（LoRA）[7] 計算由該研究提出的 Variational Score Distillation（VSD）演算法，藉由 VSD 調整3D 模型的參數，這種方法能提高3D 模型的多樣性和品質，然而其需要計算數小時才能生成出結果，雖然和我們想達到的生成速度差異過大，但其通過3D 模型渲染的2D 圖像進行最佳化的做法值得我們學習與改進。使用生成對抗網路 （Generative Adversarial Networks, GANs）[11] 的概念進行最佳化，是另一種可能的方法，藉由 Shap-E 的生成結果渲染出的多視角2D 圖像，用鑑別器（Discriminator）區分生成的2D 圖像與真實的2D 圖像讓 Shap-E 學習到如何生成更逼真的3D 模型，如同3DGEN: A GAN-based approach for generating novel 3D models from image data（3DGEN）[2] 中類似的作法，但可能發生 Wasserstein GAN [17] 提到的 Mode Collapse 問題，這會導致生成的3D 模型缺乏多樣性，此外在 Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling [14] 中，直接以鑑別器判斷3D 模型是否為生成結果，能夠不需考慮視角問題更準確的調整用於生成的潛空間向量，但在3D 計算生成器和鑑別器的多次互動和更新，會極大增加訓練時間 和資源消耗。綜上所述，Shap-E 生成結果和輸入的圖像不夠相似，而 NeRF 對於輸入的圖像要求嚴格，生成結果不便使用，因此我們使用輸入多視角圖像的特點，結合基因演算法 （Genetic Algorithm, GA）[5] ，對 Shap-E 進行最佳化，便能得到兩者的優點:計算速度快、不同視角結果一致、生成結果能用網格表示便於後續使用。

## 三、 方法與系統設計

本研究通過結合 Shap-E [10] 模型與基因演算法 （GA）[5] ，提升多張圖像生成3D 模型的與輸入圖像之相似度。如圖1所示 Shap-E 模型的圖像編碼器（image300M）[10] 圖像轉換為潛空間向量，並通過 Decoder 生成3D 模型，但因僅依靠單視角資訊，對物體背面或側面的形狀還原存在局限性。本研究利用 GA 對潛空間向量進行最佳化，藉此改善生成模型的多視角精確度與細節還原能力。

GA 作為最佳化搜尋算法，能有效最佳化 Shap-E 生成的潛空間向量，而不需修改其模型結構或耗費大量計算資源。研究設計的適應性函數基於輸入圖像與生成3D 模型的多視角渲染結果的相似度平均值計算，避免因單一視角偏差影響評估效果。

系統架構包含六個步驟：初始化種群、選擇親代、交配、突變、適應性評估與生存選擇。初始種群由 Shap-e

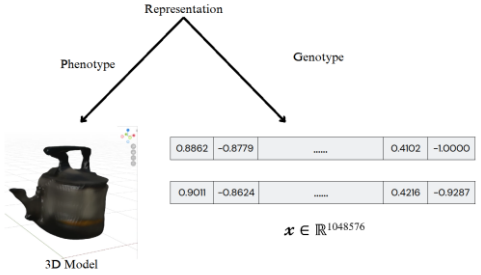


**圖1 Representation 示意圖此圖展示**

**Representation、Phenotype、Genotype 的關係和實際例子。**

針對多視角輸入，本研究提出兩種算法。一種如圖2通過計算輸入圖像的最相似視角，僅渲染關鍵視角；另一種如圖3不進行視角計算，每次完整渲染所有推測視角並計算相似度。

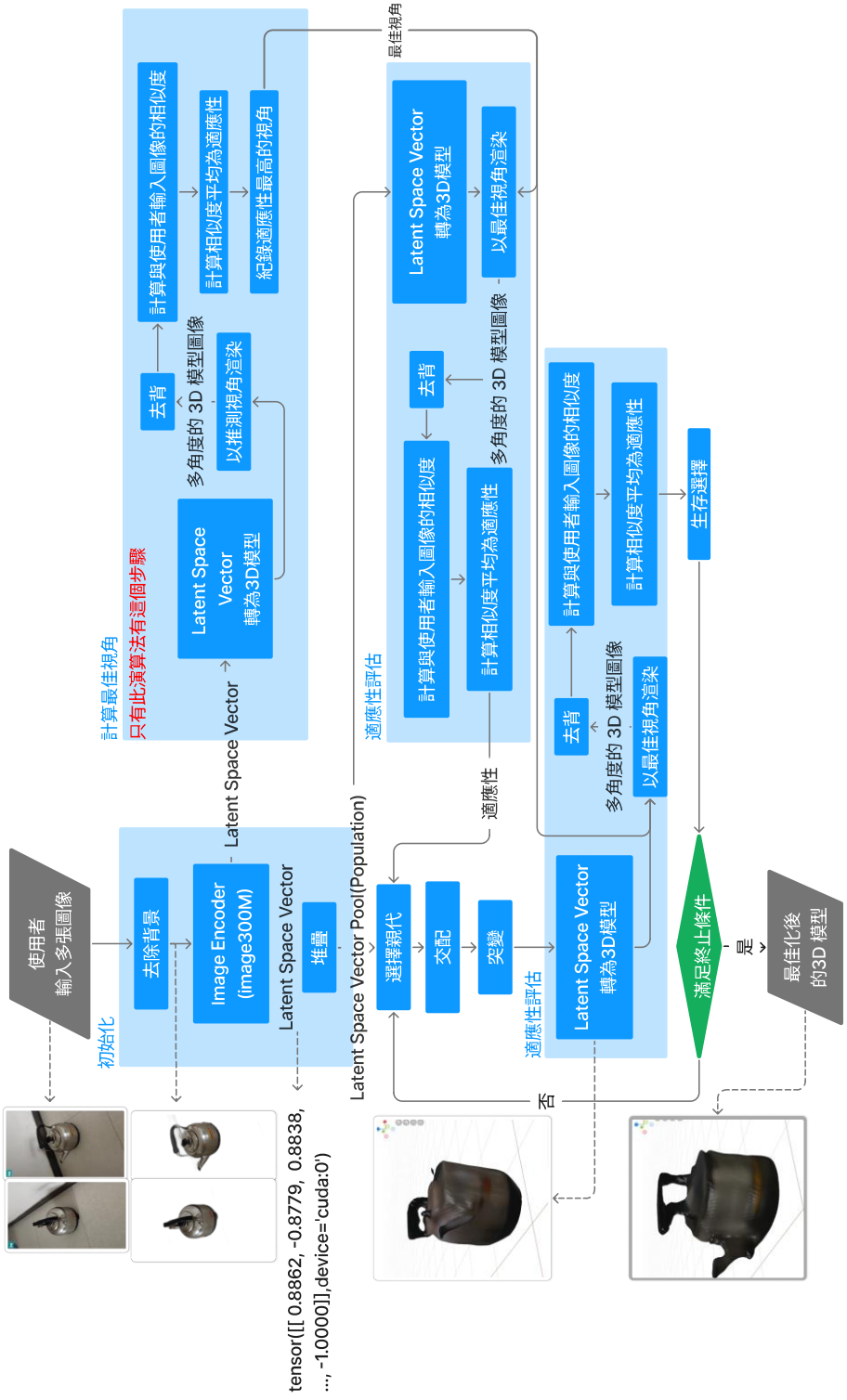

圖2 計算最佳視角流程圖此流程圖由使用者輸入圖像開始產生初始種群後，計算最接近輸入圖像的視角並經過迭代，最終得出最佳化的3D 模型。

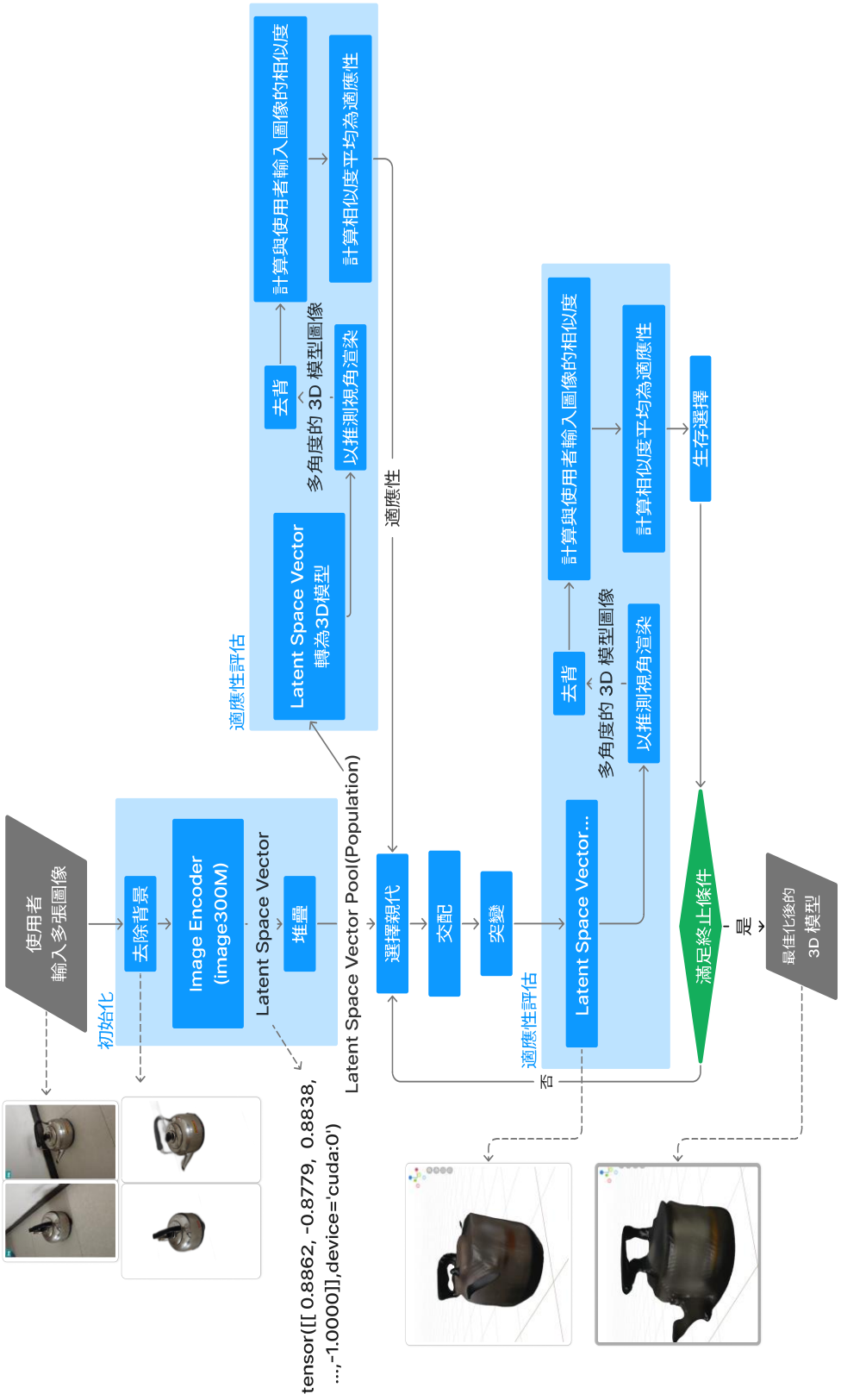


圖3 不計算最佳視角流程圖。此流程圖由使用者輸入圖像開始產生初始種群後，

**經過迭代最終得出最佳化的3D 模型。**

為展示技術成果，研究搭建了一個基於 Vue3 [8] 和 Django [6] 的網站系統，支持圖像上傳、去背處理和逐代生成結果展示，幫助用戶觀察技術最佳化過程與效果。網站整合資料管理系統，保障用戶上傳數據的隱私與安全，實現技術的可視化與應用化。

## 四、 實驗結果

在基因演算法實驗部分，我們的實驗平台為 Intel i9 13900搭配 NVIDIA RTX 4090，測試資料集為 MVImgNet [19] 及部份我們拍攝的影像。本研究實驗的參數固定為種群大小20，代表每一代的種群交配前及生存選擇後皆固定為20個染色體，固定迭代200代，推測視角為30°、120°、210°、300°，交配率為90%，即親代有90%可能交換基因、10%可能直接複製親代進行突變，突變率為0.01%，即子代每個基因有0.01%可能加上標準差0.3的常態分佈隨機數，測試資料集為物體的多視角圖像，每種資料測試隨機種子碼0~9的結果，渲染解析度固定為64X64。本研究的相似度函數計算方法採用了 VGG16 [15] 作為特徵提取工具。首先，將圖像以 VGG16模型以提取圖像特徵，接著利用餘弦相似度公式計算輸入圖像特徵與生成圖像特徵的相似程度。利用 VGG16對細節特徵的敏感性，能夠更準確地反映生成結果與目標圖像之間的差異，為後續取平均作為適應性提供可靠的基礎。


**圖4 3D 模型的輸入圖像和 Shap-e Shap-e 生成的初始3D 模型與最佳化後的3D 模型比較 。**

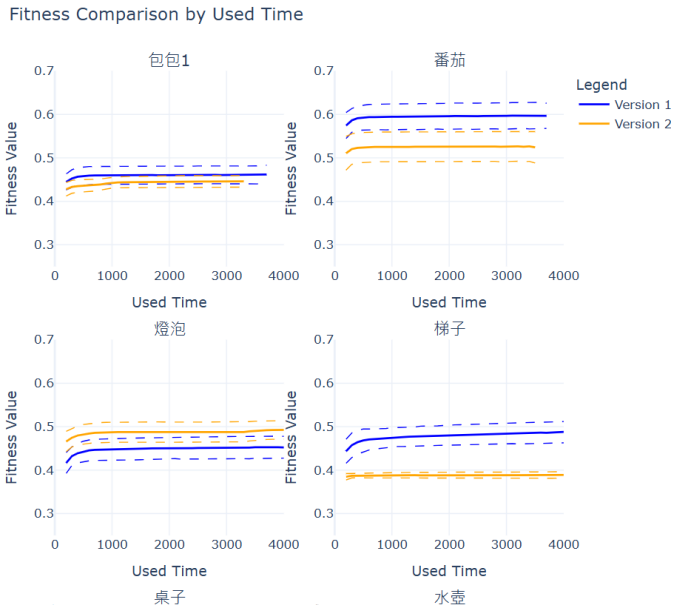


**圖5 實驗數據此圖為計算最佳視角與不計算最佳視角的結果比較。**

## 五、結論與建議

本研究在對 Shap-E [10] 生成的模型進行了改進與最佳化，並取得了有價值的研究成果如圖4。然而，實驗數據圖5顯示，仍有若干方面存在進一步提升的潛力。目前的圖像去背模型在執行效率與準確度上仍有最佳化可能。例如，探索其他更高效的模型或演算法，或許能進一步提升去背的處理速度。此外，某些特定物體的細節（如椅背與椅角）在背景去除時常被誤判為背景，這反映了模型對複雜邊緣結構的處理能力不足。在適應性函數部分，處理圖像相似度的演算法方面，現有方法對區分細微差異的能力仍顯不足。本研究成功實現了原版 Shape-E 模型所不具備的多視角生成功能，為3D 模型生成領域提 供了新的可能性。此技術的實現，不僅擴展了 Shape-E 模型的應用範圍，也為多視角生成技術的進一步研究奠定了基礎。通過引入基因演算法（GA） [5] 進行最佳化，我

們在部分背景去除較完整的圖像上取得了顯著成果。與原始 Shape-E 模型相比，經 GA 最佳化後的生成結果能更準確地還原目標圖像的細節，顯示出該方法在提高生成品質上的潛力。

我們也開發了一個基於網頁的工具，讓使用者無需經歷建立運行環境與安裝等繁瑣過程，即可輕鬆使用 GA 最佳化技術生成3D 模型。該工具顯著降低了技術門檻，為非專業用戶提供了便捷的3D 模型生成方案，具有廣泛的應用潛力。本研究為3D 圖像生成領域提供了新方法與新工具，未來我們將進一步最佳化圖像去背及適應性函數部分的模型與演算法，提升系統的精度與效率。同時，對多視角生成技術的深入研究，將探索更多創新的應用場景，如虛擬現實、遊戲設計及數位內容創作等領域。

## 六、參考文獻


[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, Attention Is All You Need, 2 Aug 2023.

[2] Antoine Schnepf, Flavian Vasile, Ugo Tanielian, 3DGEN: A GAN-based Approach for Generating Novel 3D Models from Image Data, 13 Dec 2023.

[3] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng, NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, 3 Aug 2020.

[4] Blender Online Community. (2024). Blender - a 3D Modelling and Rendering Package. Retrieved from https://www.blender.org

[5] David E. Goldberg (2002). Genetic Algorithms. Pearson Education India.

[6] Django Software Foundation. (2024, November 5). Django.

[7] Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, LoRA: Low-Rank Adaptation of Large Language Models, 16 Oct 2021.

[8] Evan You. (2024-11-15). Vue.js: The Progressive JavaScript Framework (Version 3.0). Vue.js.

[9] F. Rosenblatt, The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain,Nov 1958.

[10] Heewoo Jun, Alex Nichol, Shap-E: Generating Conditional 3D Implicit Functions, 3 May 2023.

[11] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative Adversarial Networks, 10 Jun 2014.

[12] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, Steven Lovegrove, DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 165–174, 2019.

[13] Jonathan Ho, Ajay Jain, Pieter Abbeel, Denoising Diffusion Probabilistic Models, 16 Dec 2020.

[14] Jiajun Wu, Chengkai Zhang, Tianfan Xue, William T. Freeman, Joshua B. Tenenbaum, Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling, 4 Jan 2017.

[15] Karen Simonyan, Andrew Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, last revised 10 Apr 2015.

[16] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Ronen Basri, Yaron Lipman, Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance, 25 Oct 2020.

[17] Martin Arjovsky, Soumith Chintala, Léon Bottou, Wasserstein GAN, 6 Dec 2017.

[18] Zhengyi Wang, Cheng Lu, Yikai Wang, Fan Bao, Chongxuan Li, Hang Su, Jun Zhu, ProlificDreamer: High-Fidelity and Diverse Text-to-3D Generation with Variational Score Distillation, 22 Nov 2023.

[19] Xianggang Yu, Mutian Xu, Yidan Zhang, Haolin Liu, Chongjie Ye, Yushuang Wu, Zizheng Yan, Chenming Zhu, Zhangyang Xiong, Tianyou Liang, Guanying Chen, Shuguang Cui, Xiaoguang Han, MVImgNet: A Large-scale Dataset of Multi-view Images, Submitted on 10 Mar 2023.