

NFT1000: A Cross-Modal Dataset for Non-Fungible Token Retrieval

Shuxun Wang

wangshuxun2022@ia.ac.cn

MAIS, Institute of Automation, CAS
School of AI, University of Chinese
Academy of Sciences
Beijing, China

Wei Liu

Haowei Liu

Li Yang

liuwei@ia.ac.cn

liuhaowei2019@ia.ac.cn

li.yang@nlpr.ia.ac.cn

MAIS, Institute of Automation, CAS
Beijing, China

Yunfei Lei

lyfls1998@buaa.edu.cn

Beijing University of Aeronautics and
Astronautics
Beijing, China

Bing Li*

Wenjuan Li

Jin Gao

bli@nlpr.ia.ac.cn

wenjuan.li@ia.ac.cn

jin.gao@nlpr.ia.ac.cn

MAIS, Institute of Automation, CAS
Beijing, China

Ziqi Zhang

zhangzqi2017@ia.ac.cn

MAIS, Institute of Automation, CAS
Beijing, China

Weiming Hu

wmhu@nlpr.ia.ac.cn

MAIS, Institute of Automation, CAS
School of AI, University of Chinese
Academy of Sciences
School of Information Science and
Technology, ShanghaiTech University
Beijing, China

Abstract

With the rise of "Metaverse" and "Web 3.0", Non-Fungible Token (NFT) has emerged as a kind of pivotal digital asset, garnering significant attention. By the end of March 2024, more than 1.7 billion NFTs have been minted across various blockchain platforms. To effectively locate a desired NFT, conducting searches within a vast array of NFTs is essential. The challenge in NFT retrieval is heightened due to the high degree of similarity among different NFTs, regarding regional and semantic aspects. In this paper, we will introduce a benchmark dataset named "NFT Top1000 Visual-Text Dataset" (NFT1000, as shown in Fig.1), containing 7.56 million image-text pairs, and being collected from 1000 most famous PFP¹ NFT collections² by sales volume on the Ethereum blockchain. Based on this dataset and leveraging the CLIP series of pre-trained models as our foundation, we propose the dynamic masking fine-tuning scheme. This innovative approach results in a 7.4% improvement in the top1 accuracy rate, while utilizing merely 13% of the total training data (0.79 million vs. 6.1 million). We also propose a robust metric Comprehensive Variance Index (CVI) to assess the similarity and retrieval difficulty of visual-text pairs data. The dataset will be released as an open-source resource. For more details, please refer to: <https://github.com/ShuxunoO/NFT-Net.git>

^{*}Corresponding author.

¹PFP is an abbreviation for "Profile Picture", representing a category of NFTs primarily used as avatars in social media contexts.

²An NFT collection represents an NFT project, which contains the same batch of media files and metadata data.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MM '24, October 28–November 1, 2024, Melbourne, VIC, Australia

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0686-8/24/10
<https://doi.org/10.1145/3664647.3680903>

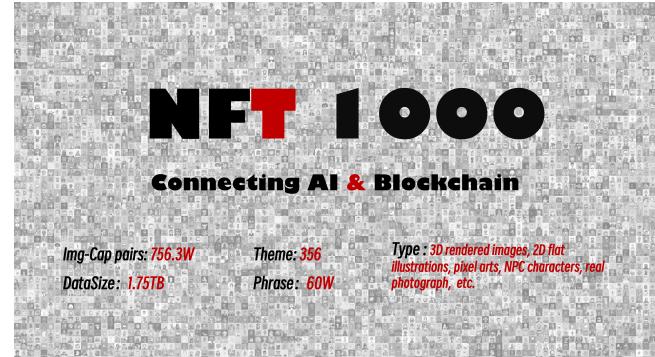


Figure 1: NFT1000 is the first NFT dataset within the field of computer vision. The proposed dataset encompasses the most renowned 1,000 avatar-based NFT projects on the Ethereum mainnet, comprising 7.563 million image-text pairs.

CCS Concepts

• Computing methodologies → *Image representations*.

Keywords

Cross-Modal Retrieval, Blockchain, NFT, CLIP, AIGC

ACM Reference Format:

Shuxun Wang, Yunfei Lei, Ziqi Zhang, Wei Liu, Haowei Liu, Li Yang, Bing Li, Wenjuan Li, Jin Gao, and Weiming Hu. 2024. NFT1000: A Cross-Modal Dataset for Non-Fungible Token Retrieval. In *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24), October 28–November 1, 2024, Melbourne, VIC, Australia*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3664647.3680903>

1 Introduction

With the emerging concept of the "Metaverse" [16, 26] and "Web3.0" [13], NFT [25] has entered the public eye as a significant digital asset

within this space. The NFT, standing for Non-Fungible Token, is a unique cryptocurrency token on blockchain [29] representing digital assets such as images, videos, tickets, inscription, etc. NFT is coveted for its characteristics of provenance, high liquidity, and rarity. NFT possesses immense value; for instance, the renowned NFT project CryptoPunks has amassed a trading volume of \$2.78 billion since its launch³. Statistical data⁴ indicates that by the end of March 2024, the cumulative number of NFT minted on different blockchain platforms has exceeded 1.7 billion. When purchasing NFTs, people often gravitate towards tokens that align with their personal style or match their preferences, aiming to fulfill their desire for personalized expression in virtual spaces. However, both the academic and industrial sectors lack effective and precise methods or toolkits for the retrieval of NFT data due to the high degree of regional and semantic similarity among NFTs (Fig.2). This represents a novel research area that requires our exploration.

Given the lack of a dedicated NFT dataset for scientific research in the field of computer vision, we firstly construct the NFT1000. It is composed of the top 1000 PFP NFT collections by sales volume on the Ethereum blockchain with the ERC-721⁵ standard. Each project contains an average of 7500 image-text pairs. In total, the dataset includes 7.56 million image-text pairs, with a data volume of 1.75TB. It is suitable for various downstream tasks such as retrieval, generation and so on.

Under the background of NFT-type data retrieval and leveraging the NFT1000 dataset, we introduce a task focused on large-scale, high-similarity image-text retrieval, representing a potential approach in the intersection of AI and blockchain research. This task aims to retrieve target images from a massive collections of highly similar pictures by using tokens' descriptions. Although CLIP models are pre-trained using 400 million image-text pairs from the Internet, their performance on fine-grained classification tasks is somewhat lacking. This indicates that CLIP's training approach struggles to capture the local semantic information of image-text pairs. To address the limitation, we propose a dynamic masking fine-grained contrastive learning scheme. Through analysis of input images, its dynamic masking module probabilistically masks certain component areas of the image and the corresponding captions. This subtractive approach from the global semantics more fully exposes the local features of the image-text pairs, allowing the model to more specifically align the detailed information of the visual-caption pairs. Our experimental results demonstrate that it is possible to train a model that surpasses the total data's top1 accuracy by 7.4% using only 13% of its training data. This significantly reduces the training overhead and enhances the effectiveness of data utilization.

To quantitatively assess the similarity between a set of images and texts, rather than relying on subjective human judgment, we propose the Comprehensive Variance Index(CVI). It comprehensively considers the similarity within images, captions, and the degree of match between images and texts. Our empirical evidence

³As of April 12, 2024, 22:00, the data is sourced from site of <https://nftgo.io/macro/market-overview>

⁴<https://www.nftscan.com/>

⁵ERC-721 stands for Ethereum Request for Comment #721. It is a universal NFT standard protocol that defines a series of interfaces for NFT token transactions. For more details, please visit: <https://eips.ethereum.org/EIPS/eip-721>.

demonstrates a clear correlation between this index and retrieval accuracy.

Our main contributions are: (1) We construct the first NFT visual-text dataset in the field of computer vision. (2) We introduce a task of large-scale, high-similarity image-text retrieval. (3) We design an effective training method for NFT data, using less data but training better models. (4) We propose the Comprehensive Variance Index, a universal metric designed to measure the similarity between images and texts.



Figure 2: Randomly selecting seven projects and choosing seven images from each to create an average image (as shown in the red-framed picture), we can observe that the average image has clear contours and distinct content. This indicates that the batch of images randomly selected from the same project possesses a high degree of regional similarity.

2 Related Work

2.1 About NFT

NFT [25], short for Non-Fungible Token, is a kind of unique virtual digital asset based on blockchain [29]. As a fundamental component of the metaverse, NFT plays a significant role in various domains, such as social interaction, finance, sports, gaming copyright verification, etc. NFT is a broad concept encompassing a diverse array of forms, including images, videos, text, audio, code, and more. Each form of NFT is unique, making them distinct from more common, interchangeable tokens like cryptocurrencies (e.g. Bitcoin [17] and Ethereum [1]). However, the most widely accepted forms of NFT currently are multimedia formats such as images and videos.

NFT is highly valued due to its unique combination of scarcity, verifiability, liquidity, and the ability to fulfill people's social status needs. NFT is scarce because each one is unique or limited in quantity, making it sought after in a market where people are willing to pay more for rare items. Its possession is verifiable through blockchain technology, which provides a secure, transparent record of each NFT's history and ownership, ensuring authenticity and reducing the risk of fraud. Furthermore, NFT offers high liquidity compared to physical assets; it can be easily bought, sold, or traded on global platforms with minimal transaction costs, making it attractive to investors looking for quick and efficient asset turnover. Lastly, owning an NFT, especially those created by famous artists or those that are particularly rare, can convey social status, as it signifies wealth, taste, and exclusivity. This desire for social recognition through unique digital assets drives demand and increases its value. Collectively, these factors make NFT valuable in today's digital economy, appealing to collectors, investors, and those seeking social distinction alike. According to statistical data⁶, prominent

⁶As of April 12, 2024, <https://nftgo.io/discover/top-collections>

NFT projects have achieved significant trading volumes: Bored Ape Yacht Club has sold \$3.66 billion, CryptoPunks has amassed \$2.78 billion, Mutant Ape Yacht Club has made \$2.51 billion, etc.

As the metaverse continues to develop, NFT will increasingly become a digital commodity for trading. As previously mentioned, an NFT can significantly represent the taste of its holder. Therefore, consumers often prefer those that are renowned and align with their personal style. However, with billions of NFT entries, finding one that suits an individual's needs is challenging. Additionally, the high degree of similarity among NFTs adds considerable complexity to their retrieval. Thus, the task of retrieving an NFT is both a critical need and highly challenging, meriting in-depth research and exploration.

2.2 Cross-Modal Retrieval

Cross-modal Image-text Retrieval (ITR) is to retrieve the relevant samples from one modality while the queries are expressed in another modality, usually consists of two subtasks: image-to-text (i2t) and text-to-image (t2i). ITR has been witnessed great success in recent years [4, 11, 19] thanks to the rapid development of deep language-vision models [3, 5, 23] and various large-scale multi-modal pre-trained models [8–11, 19, 21, 27]. Most ITR systems deployed in real-world applications are built upon pre-trained models that have been fine-tuned. Generally speaking, the pre-trained models can be divided into two categories according to the their architectures: 1) Fusion-structure models: ALBEF [11] and BLIP [10]. 2) Dual-encoder models: CLIP [19], META-CLIP [27] and VILEM [2]. The fusion-structure models process text and image inputs simultaneously through a unified network architecture. In these models, image and text data are merged at an early stage and the entire model propagates forward through a single data stream. The drawback of fusion-structure models is their low-efficiency and inflexibility due to the computation of similarity between queries and whole data of another modality during retrieval. While dual-encoder models encode image and text in parallel by independent models and align them by self-supervised contrastive learning. Compared with fusion-structure models, dual-encoder models are more flexible and are much more efficient at zero-shot inference [19]. Dual-encoder models align image and text semantic features into a consistent high-dimensional feature space and the encoders are generally pre-trained models. Besides, the computed semantic features of each branch can be stored for fast inferring during retrieval. These advantages make dual-encoder models efficient and flexible to deploy. In this work, we will fine-tune a series of dual-encoder models on our NFT1000 dataset.

2.3 Image-Text Dataset

In the realm of computer vision and natural language processing, datasets like Flickr30K [18], COCO [12] and LAION-5B [20] offer vast amount of image-text pairs for diverse applications. Flickr30K is an image-caption dataset widely used in computer vision and natural language processing research. It consists of 31,000 images sourced from the online photo-sharing platform Flickr. Each image in the dataset is paired with five English captions, which provide descriptive annotations written by human annotators. The COCO dataset provides over 200,000 labeled images with detailed instance

annotations and The LAION-5B encompasses 5.85 billion CLIP-filtered image-text pairs, making the training of large-scale multi-modal models plausible.

However, Most of the data in the above datasets are collected from the real world, which inherently exhibits significant distributional differences compared to NFT data. In addition to this, images from one NFT project, although different, have fine-grained semantic similarity because they are permutations and combinations of fixed components, as we will discuss in Section 3.2, this is a distinctive feature that the aforementioned datasets do not possess. To our knowledge, iCartoonFace benchmark [28] has similar situation with NFT1000, it is a large-scale, high-quality, richly annotated cartoon face recognition dataset, containing 389,678 images of 5,013 cartoon characters. However, this dataset lacks captions corresponding to each image, making it difficult to meet the requirements for cross-modal retrieval.

Given the absence of a dedicated NFT dataset in the computer vision field, in this work, we construct the first benchmark dataset consisting of NFTs, designed to support NFT retrieval and generation tasks.

3 Properties of NFT1000

3.1 Inherent Image-Text Pair Format

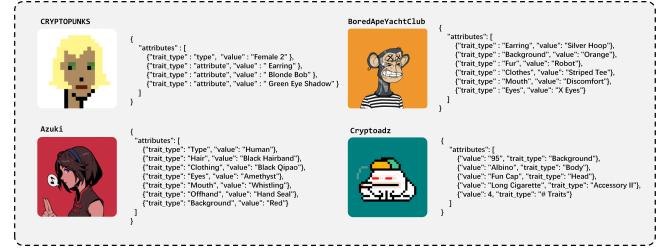


Figure 3: In the NFT1000 dataset, each image within every collection naturally comes with an accompanying JSON file, which introduces the attributes of the image in a key-value pair format.

Each NFT in the dataset is associated with a metadata resource file, which typically exists in the form of a JavaScript Object Notation (JSON) format. This file uses key-value pairs to describe the attributes of the NFT token(Fig.3).

3.2 Fixed-Components Permutation and Combination

The essential reason for the high degree of similarity among NFT images within a same project lies in the fact that all images are permutations and combinations of fixed components. As shown in Fig.4: (a) Images contain a clothing layer named “Navy striped tee”; (b) Pictures include the same “3D glasses” layer. (c) Every image features the same “Bored bubblegum mouth”. (d) All photos are adorned with a same “Commie hat”. However, it is important to note that in projects initiated after the removal of identical image covers, no two images within a project are the same.

3.3 Abstract Description

NFT can be considered a form of crypto arts, but the definition of these artworks by artists often include subjective elements. This

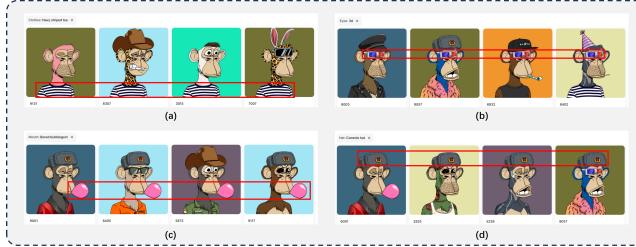


Figure 4: All images within the same collection are blended from a specific set of components arranged in various combinations, resulting in pixel-level uniformity in image regions.

leads to the abstract description issue, which can be understood as the image itself being difficult to comprehend or the image description lacking clear semantic information. From Fig.5, we can observe intuitively that the No.0 token from the *Superlative Secret Society* project is particularly hard to comprehend, or rather, there is no obvious correlation between its image and caption. It is noteworthy that this situation is common in NFT projects.

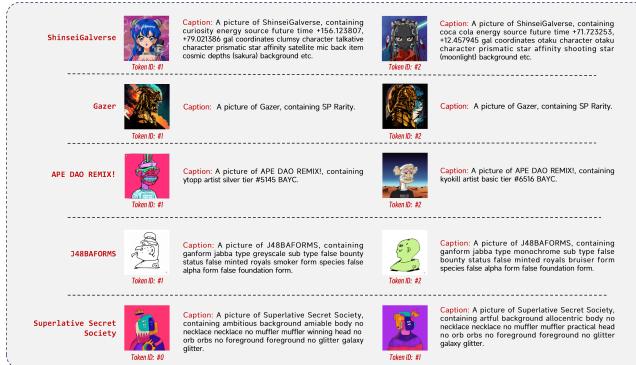


Figure 5: Show case of abstract images and their abstract descriptions.

4 Constructing NFT1000

4.1 Clarifying the Download Targets

Among various NFT categories, PFP NFT collections account for over 60% of the market share⁷. Besides, in avatar-type NFT, the JSON file accompanying each image relatively effectively describes its own attributes. Ethereum is the birthplace of NFT and the most flourishing blockchain for NFT crypto arts. Therefore, we select the top 1000 PFP NFT projects on the Ethereum blockchain, based on sales volume, as our download targets.

4.2 Downloading and Filtering

We utilize resources from the Web3.0 domain such as NFTScan⁸, Alchemy⁹ and IPFS¹⁰, leveraging the basic resource links provided

⁷Please refer to the “Category Market Cap” entry on the Web: <https://nftgo.io/analytics/market-overview>

⁸<https://www.nftscan.com/>

⁹<https://www.alchemy.com/>

¹⁰<https://ipfs.tech/>

in the smart contracts¹¹ of each NFT project. This enabled us to piece together the complete links for the media resource and JSON data of each token for downloading and collection. In fact, we have downloaded resources from a total of 1250 projects for purpose of selection.

Among all the collections that have been fully downloaded, we exclude those with completely duplicated media data (or all images being identical covers), projects with an insufficient total number of tokens (set as fewer than 500), and those lacking a JSON file or where the JSON file contains no substantive semantic information.

4.3 Standardization

Standardize File Format and Dimensions. Native NFT data, encompassing static image formats such as JPG, PNG, SVG and WebP, are uniformly transformed into the PNG format (This conversion is primarily due to PNG being the predominant format in most NFT collections, and the choice is intended to maximally retain the original fidelity of the data). For dynamic media formats, including GIF and MP4, a representative frame is randomly selected and converted into PNG format. The standardized resolution for these images is set to a width of 512 pixels, with a proportionally adaptive height to maintain aspect ratio integrity. Employing this method, we have reduced the original data size from 14TB to 1.75TB.

Caption Extraction. For the original key-value pairs formatted attribute lists, there are two methods for generating captions(Fig.6): one is based on large language models (ChatGPT, LLAMA-13B [22]), using prompt engineering to create descriptions according to the attribute list corresponding to the image; the other way involves using predefined sentence templates to concatenate attributes into a single caption. By using large language models, we generate 30,000 descriptions for 10,000 randomly selected images, while also creating 10,000 captions using language templates. Subsequently, we utilize OpenAI’s CLIP-ViT-L pretrained model for zero-shot inference and compared the retrieval accuracy of captions obtained via the two methods (Fig. 7). The result indicates that the large language model can generate better image descriptions, but overall, the performance of the two methods **does not differ significantly**. Lastly, considering the former method would consume considerable time and computational resources, we ultimately opt for generating captions using sentence templates.

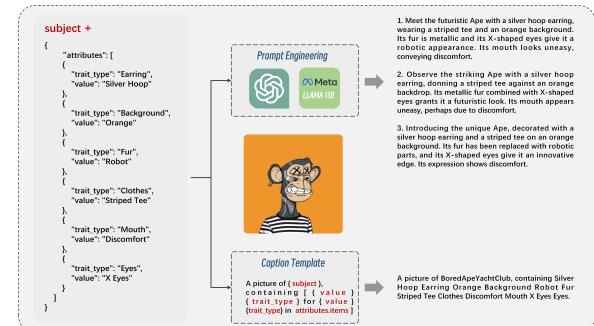


Figure 6: Illustration of two methods for generating captions

¹¹Smart contracts on blockchain are self-executing scripts with the terms written in code.

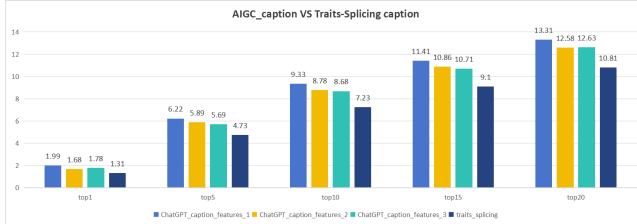


Figure 7: The CLIP model’s zero-shot retrieval accuracy in comparing captions generated by large language models vs. those produced by sentence template.

Data Partitioning. Due to the presence of identical components and descriptions in images within the same project, internal division of training and test sets in a NFT collection may result in “data leakage.” Consequently, we adopt the project as the fundamental unit for data division, allocating the entire dataset into training, validation, and test sets in an 80:5:15 ratio.

Dataset Statistics. The NFT1000 dataset comprises 1000 outstanding PFP NFT projects, each containing approximately 7500 image-text pairs, encompassing a total of 7.56 million image-text pairs with a collective data volume of 1.75TB. In the dataset, the training set includes 800 projects with 6,178,249 image-text pairs. The validation set comprises 50 projects with 383,916 image-text pairs, and the test set consists of 150 projects with 1,000,838 image-text pairs. The content spans a diverse range of artistic types, including 3D rendered images, 2D flat illustrations, pixel arts, NPC characters, real photographs, etc. It covers a total of 356 different content themes and 595,504 unique descriptive phrases. For more details about each NFT project, please refer to Appendix Table.

5 Fine-Grained Contrastive Learning

The CLIP models gain fame for achieving state-of-the-art (SOTA) performance through zero-shot inference on various datasets, following its training on a dataset of 400 million image-text pairs using a straightforward contrastive learning strategy. We sequentially use OpenAI’s CLIP-ViT-B-32, CLIP-ViT-L-14 pretrained models and META’s META-CLIP-ViT-L-14 [27] for zero-shot inference and fine-tuning. The experimental results are presented in Table 1. This table reveals that these pretrained SOTA models have almost never encountered data from the NFT1000 dataset, indicating that the data distribution in NFT1000 is unique and novel. Despite the noticeable improvement (with an average increase in top1 accuracy of about 10%), the overall effectiveness remains suboptimal.

Table 1: Comparison of zero-shot inference and fine-tuning inference accuracy of different models on the NFT1000 test set.

model-type	zero-shot			fine-tuning		
	top1	top5	top10	top1	top5	top10
CLIP-ViT-B-32	0.01	0.02	0.03	10.63	20.32	25.19
META-CLIP-ViT-L-14	0.00	0.01	0.02	13.06	23.68	28.81
CLIP-ViT-L-14	0.06	0.25	0.42	15.36	27.55	33.26

As discussed in Section 3.2, all images within an NFT project are permutations and combinations of fixed components. Given

that the CLIP model is not particularly adept at focusing on the local semantic information of images, we hypothesize that the prerequisite for precise retrieval is accurate cognition. If we could fine-tune the CLIP model at the component level, it might address the issue of the fine-tuned model not achieving satisfactory recall performance. To verify this hypothesis, we propose a fine-grained fine-tuning strategy based on dynamic masking.

5.1 Component Separation



Figure 8: Illustration of component separation. The results demonstrate that, through a process of initial differentiation followed by superposition, components can be separated into relatively clean and complete entities, even in the presence of overlapping among them.

Given the pixel-level consistency within the same area of images containing the same component in an NFT project, we adopt a strategy of differentiation followed by superposition to isolate the various distinct components. The specific approach is as follows:

- (1) Identify which images share a same component, achievable through analysis of the NFT’s accompanying JSON file.
- (2) Randomly select a set of images, using the first image as a template, and perform image differencing operations with the subsequent images to get the shared regions and their mask representations.
- (3) Repeat step 2 multiple times, ultimately assembling the fragmented components into a relatively complete component and its mask.

Experiments show that performing differencing operations on 4 images at one time and repeating this process 8 times is a good choice. This combination balances execution efficiency and also results in relatively complete and clean components and masks, as shown in Fig. 8.

5.2 Dynamic Masking

Before the model loads the training image-text pairs data, we firstly analyze the image to identify its constituent components. With probability p , a component’s corresponding mask is randomly selected to perform a masking operation on the original image. Simultaneously, the tag of the selected component is removed from the full caption. This process results in a new image-text pair that lacks certain local pixels and descriptive information, thereby allowing the detailed information of the image-text pair to emerge from the global semantics. By subtracting from the original image-text pairs in this manner, the model is encouraged to fully comprehend

the correspondence between components and their names, thereby achieving fine-grained feature alignment with NFT data. A dynamic visualization of the masking process is shown in the Fig. 9.

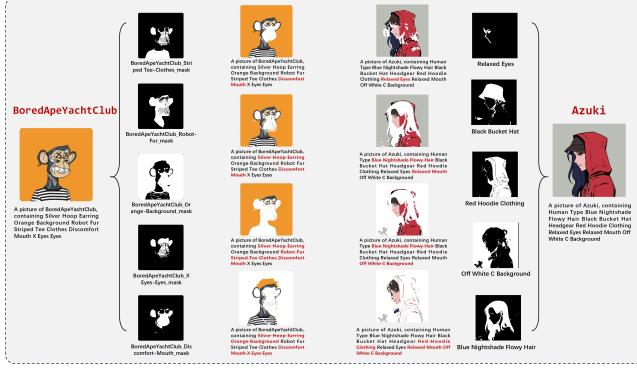


Figure 9: Illustration of the generation process of dynamic masks. Through this method, a single NFT image-text pair can generate new pairs with varied semantic richness.

6 Experiments On NFT1000

In this chapter, we will conduct a series of experiments to validate the effectiveness of the dynamic masking fine-tuning method and the application potential of the NFT1000 dataset, focusing on four aspects: the selection of the dynamic masking probability p , the generalizability of the dynamic masking approach, the metric Comprehensive Variance Index (CVI) and NFT generation.

In the classical contrastive learning framework, we introduce a dynamic masking unit capable of analyzing the composition of sampled image-text information. This unit applies masks to components of an NFT image with a specific probability, thereby eliminating certain semantic information from the global image-text context. For the remaining training pipeline, we employ the same training strategy as the original CLIP to fine-tune models. Specifically, we utilize image and text encoders to extract features from images and captions. Subsequently, we use contrastive loss to optimize the parameters of the image and text encoders, aiming to progressively align NFT images and their corresponding captions within the same semantic space. The training pipeline is illustrated in Fig. 10.

6.1 Mask Selection Probability

During the process of generating dynamic mask, a component mask is selected with a probability p . The larger the value of p , the more areas of the original image are masked, resulting in finer semantic granularity but also a more fragmented image; conversely, the smaller the value of p , the fewer areas are masked, leading to coarser semantic granularity and a more rough correspondence between components and captions. Therefore, selecting an appropriate p is a critical issue.

To swiftly determine the appropriate probability, we construct a smaller dataset from the complete dataset, called NFT1000mini. This subset consists of a training set with 800 projects, a batch of 1000 image-text pairs are randomly extracted from per project, totaling 794,698 pairs; and a test set comprising 150 projects, each with 1000

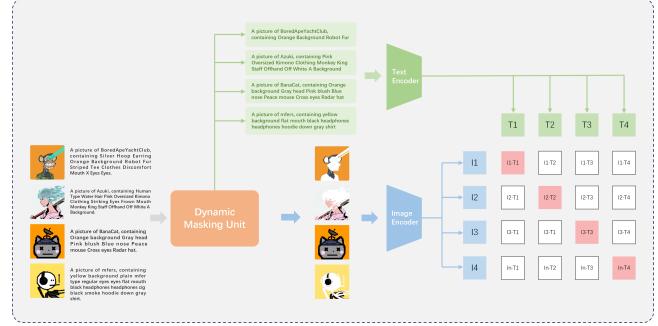


Figure 10: Illustration of the fine-tuning pipeline. The integration of the dynamic masking unit allows for the highlighting of local information within NFT image-text pairs, thereby facilitating fine-grained alignment of the model with NFT data.

Table 2: Comparison of data sizes between NFT1000mini and NFT1000

	NFT1000mini		NFT1000	
	NFT project number	image-text pairs	NFT project number	image-text pairs
training set	800	794,698	800	6,178,249
validation set	50	49,738	50	383,916
test set	150	147,615	150	1,000,838

random image-text pairs, totaling 147,615 pairs. The comparison of data sizes between NFT1000mini and NFT1000 is shown on Table 2. Subsequently, we conducted a series ablation studies by using the pre-trained CLIP-ViT-B-32 model on the NFT1000mini training set with the same training parameters but varying p for model fine-tuning. With results shown in Table 3 . From the table, we can observe that the relationship between p and accuracy forms a convex function, peaking near $p = 0.5$. This indirectly suggests that the more random the mask selection, the better the training effect of the model. Unless otherwise specified, we set $p = 0.5$ in subsequent experiments.

Table 3: The impact of different mask selection probabilities on model retrieval performance.

probability	top1	top5	top10
p=0	16.93	29.99	36.17
p=0.3	22.79	36.86	42.87
p=0.5	22.68	37.17	43.51
p=0.7	21.59	35.71	41.88

6.2 Generalizability of Dynamic Masking

To verify whether we can fine-tune the model more efficiently under the condition of fine-grained semantic alignment, we compared the inference performance on the NFT1000mini test set of different models trained with and without dynamic masking on the NFT1000mini training set, as well as those trained on the entire NFT1000 training dataset. Subsequently, we obtained surprising results, as shown in Table 4. It is evident that under the same training set conditions (NFT1000mini training set), the use of dynamic masking leads to at least a 10% improvement in accuracy. Compared

with the CLIP-ViT-L-14 model, which achieves SOTA performance using the NFT1000 training set, there's a 7.44% increase in top1 accuracy. This conclusively demonstrates the effectiveness of the dynamic masking training method.

In addition to conducting instance-level searches across the entire dataset, we also compared the search results within a specific NFT project by zero-shot and fine-tuning inference, with data presented in Table 5. It displays the retrieval results for the top 5 and bottom 5 NFT projects, with the data for the top 5 achieving nearly 100% in the top 10 accuracy. However, we can also directly observe that the bottom 5 NFT projects show almost no improvement in accuracy before and after model fine-tuning. This issue arises from the abstract definitions discussed in section 3.3. Consequently, how to retrieve NFT data with abstract definitions will be a focal point of our future work.

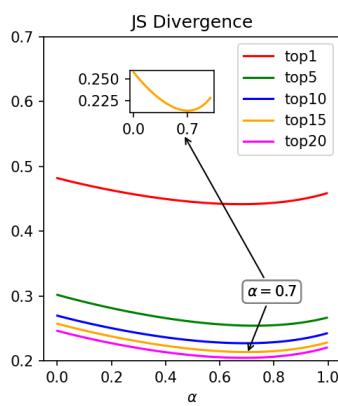


Figure 11: After L1 normalization, the trend of the JSD between the CVI distribution and the TopK distribution varies with changes in alpha. When α approaches 0.7, the JSD approximately reaches its lowest point and CVI can most accurately serve as a measure of image-text similarity.

6.3 Comprehensive Variance Index

To quantitatively measure the similarity between a set of image-text pairs, rather than just relying on subjective human judgment (for example: "not very similar," "somewhat similar," "very similar", etc.), we propose the Comprehensive Variance Index. In the current realm of deep learning, a commonly used approach [14] for image-text retrieval involves employing pretrained visual and language encoders to extract image and text features, known as embeddings. Subsequently, dot product operations are conducted to obtain the cosine similarity between the images and texts. The similarity scores are then sorted in descending order to yield the final topk results. For any given model, the most hard retrieval scenario occurs when all probabilities are identical, forcing the model to make a blind selection.

Based on this observation, we propose a concept originating from the probability distribution of vector cosine similarities. This concept posits that if a batch of images exhibits a more uniform distribution of cosine similarity probabilities (in a certain sense, the smaller the variance in the distribution of cosine similarity),

the features of these images are more similar. This similarity manifests in semantic and regional aspects of the images, concurrently increasing the difficulty of image retrieval.

Drawing from the preceding discussion, we propose the Comprehensive Variance Index. $I \in \mathbb{R}^{N \times M}$ represents the feature vectors of a batch of images, in which N represents the number of images and M denotes the dimensionality of the feature vectors. Then $S_{II} \in \mathbb{R}^{N \times N}$ is given by $S_{II} = I \cdot I^\top$. Similarly, we can obtain the inner product of the corresponding texts' feature vectors, denoted as $S_{TT} \in \mathbb{R}^{N \times N}$, and the inner product of the text-image feature vectors, denoted as $S_{TI} \in \mathbb{R}^{N \times N}$. Following, CVI of a batch of image-text pairs is defined as

$$\text{CVI} = \frac{1}{2N} \left(\alpha \sum_{i=1}^N \text{var}(S_{II_i}) + (1 - \alpha) \sum_{i=1}^N \text{var}(S_{TT_i}) + \sum_{i=1}^N \text{var}(S_{TI_i}) \right) \quad (1)$$

where i represents a row in the matrix, α stands for the bias index, indicating the overall metric's preference for the similarity between images and the similarity between captions.

Jensen-Shannon divergence (JSD) [15] is a popular method for measuring the similarity between two probability distributions. It is a symmetrized and smoothed version of the Kullback-Leibler divergence (KLD) [24]. Given two probability distributions P and Q , the JSD is mathematically defined as:

$$JSD(P \parallel Q) = \frac{1}{2} D(P \parallel M) + \frac{1}{2} D(Q \parallel M) \quad (2)$$

where $M = \frac{1}{2}(P + Q)$. One of the key properties of JSD is its boundedness, as it ranges from 0 to 1. A value of 0 indicates that the two distributions are identical, while a value of 1 signifies complete dissimilarity.

Experiments show that when α is approximately 0.7 (Fig.11), CVI best fits the experimental data. This also suggests that the information contained in images is more significant than that in captions and should therefore have a greater weight in similarity measurements. We randomly selected some projects from NFT1000 and some categories from COCO [12] to conduct zero-shot inference using a pretrained CLIP model and to calculate the corresponding CVI values. The results are shown in Table 6, we can see that the lower CVI value, the more similar the batch of image-text pairs is, indicating a higher retrieval difficulty; conversely, a higher CVI value signifies easier retrieval. This also demonstrates that data retrieval within the NFT1000 dataset is indeed a challenging task.

6.4 Applications of the dataset

We developed a NFT retrieval system based on the models we have trained and use it for piracy detection task for NFTs, illustrated in Fig.12(a), which aids in copyright protection for well-known NFT projects.

As discussed in Section 3.1, NFT data inherently comes with a descriptive JSON file, and most NFTs fall within the category of artworks, making them particularly suitable for generative tasks. Leveraging diffusion models [6] and LoRA [7], we also trained several LoRA plugin models for the Azuki Fig.12(b) and Akutars

Table 4: Inference results of different models on the NFT1000mini test set under various training methods.

model_type	zero-shot			FT-NFT1000mini			FT-NFT1000			FT-NFT1000mini-with-dynamic-mask		
	top1	top5	top10	top1	top5	top10	top1	top5	top10	top1	top5	top10
CLIP-ViT-B-32	0.03	0.10	0.15	12.10	23.66	29.54	20.33	34.47	40.74	22.68 ↑ 2.35	37.17 ↑ 2.7	43.51 ↑ 2.77
META-CLIP-ViT-L-14	0.01	0.05	0.11	20.53	35.05	41.67	23.07	37.08	43.01	31.83 ↑ 8.76	47.29 ↑ 10.21	53.47 ↑ 10.46
CLIP-ViT-L-14	0.33	1.02	1.55	20.43	34.78	41.13	26.66	41.91	48.22	34.10 ↑ 7.44	50.21 ↑ 8.3	56.41 ↑ 8.19

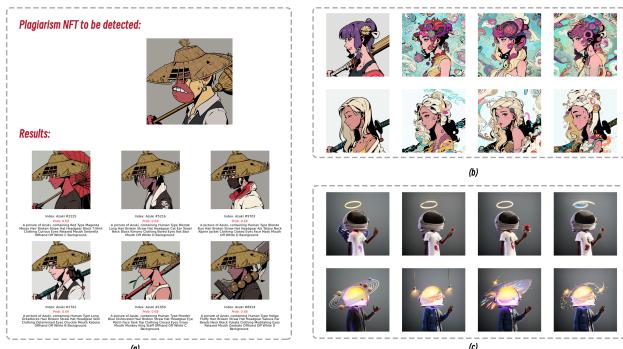
Table 5: The recall rate within NFT project before and after fine-tuning the CLIP-ViT-L-14 model.

collection	item_num	zero-shot			fine-tuning		
		top1	top5	top10	top1	top5	top10
Stoner Ape Club	6666	1.08	2.81	4.29	91.31	98.93	99.61
Junglebayapeclub	5555	0.90	2.65	4.14	89.79	98.56	99.23
Cool Ape Club	5555	0.59	1.39	2.32	88.17	97.95	99.05
Fat Rat Mafia	7777	0.03	0.27	0.63	83.27	96.18	98.06
0xAzuki	9999	0.85	3.25	5.48	80.73	95.44	97.82
.....
ShinseiGalverse	8889	0.01	0.16	0.35	0.19	0.75	1.31
Gazer	2100	0.00	0.24	0.43	0.05	0.19	0.43
APE DAO REMIX!	5528	0.02	0.14	0.22	0.04	0.18	0.36
J48BAFORMS	4848	0.04	0.10	0.21	0.12	0.27	0.54
Superlative Secret Society	11110	0.02	0.05	0.08	0.02	0.08	0.21
all_collections	1000838	0.06	0.25	0.42	21.04	34.40	40.16

Table 6: Comparison table of retrieval accuracy and CVI between NFT1000 and COCO datasets.

	Collection/Category	Top1	Top5	Top10	CVI
NFT1000	Savage Droids	0.0365	0.1823	0.3646	0.0003
	Horizon	0.0286	0.1429	0.4858	0.0005
	CyberTurtles	0.3060	0.7201	1.3141	0.0007
	SpriteClub	0.4758	1.7359	2.9574	0.0009
	Tasty Bones	1.4656	3.7433	5.9418	0.0011
COCO	person	10.3192	25.6125	37.2309	0.0034
	car	13.2463	36.3806	50.0000	0.0037
	broccoli	18.0556	37.5000	56.9444	0.0038
	backpack	24.8908	52.8384	68.1223	0.0042
	cell phone	26.0465	50.2326	60.4651	0.0044

Fig.12(c) NFT projects. They demonstrate the capabilities of generative models for the open-ended creation of NFTs with varying styles and the editable generation of NFTs within the same style.

**Figure 12: The potential applications based on the NFT1000**

7 Discussion and Future Work

7.1 Efficient Utilization of Data

As shown in Table 4, using just 13% of the training dataset, we trained a superior model. This raises the question: What is the minimum data needed for accuracy? Efficient data use requires further exploration, as does retrieving NFT projects with abstract definitions, as discussed in Section 3.3.

7.2 Continuing to Expand the Dataset

NFT1000 is an ambitious project. In the future, we plan to broaden our scope beyond Ethereum to include more collections of outstanding NFTs from other public blockchains like Solana, Polygon, BNB Chain, Klaytn, etc. We aim to scale the data to the level of hundreds of millions, striving to build an ImageNet equivalent in the NFT domain, thereby making a significant contribution to both the academic and industrial communities.

7.3 Exploring Further Potential of NFT1000

NFT holds significant untapped potential for development. In the future, we plan to explore the use of generative models to create a wider array of NFT artworks.

8 Conclusion

In this work, we construct the first NFT visual-text dataset in the field of computer vision. Furthermore, we propose an effective training method for NFT-type data, called dynamic masking fine-tuning scheme, and have trained several models as our baseline. To quantify image-text similarity, we introduce the Comprehensive Variance Index, which accounts for the similarities within images and texts, as well as the degree of image-text matching. Finally, we also explore the application of NFT data in the image generation field.

Acknowledgments

This work was supported by the Key Research and Development Program of Xinjiang Urumqi Autonomous Region under Grant No. 2023B01005, the National Natural Science Foundation of China (Grant Nos. 62302501, 62036011, 62122086, 62192782, 61721004, 62202469, 62066011 and U2033210) as well as CCF-Tencent Rhino-Bird Open Research Fund. Bing Li is also supported by Youth Innovation Promotion Association, CAS.

We would also like to thank our partners: WTF Academy, NFTScan, NFTGo, Alchemy, OpenSea, GCC, 0xAA, Quan Yuan, Yabo Li, Boyu Cai, and others who provided valuable assistance in the research and preparation of this work.

References

- [1] Vitalik Buterin et al. 2014. A next-generation smart contract and decentralized application platform. *white paper* 3, 37 (2014), 2–1.
- [2] Yuxin Chen, Zongyang Ma, Ziqi Zhang, Zhongang Qi, Chunfeng Yuan, Ying Shan, Bing Li, Weiming Hu, Xiaohu Qie, and JianPing Wu. 2023. ViLEM: Visual-Language Error Modeling for Image-Text Retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11018–11027.
- [3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
- [4] Andrea Frome, Greg S Corrado, Jon Shlens, Samy Bengio, Jeff Dean, Marc'Aurelio Ranzato, and Tomas Mikolov. 2013. Devise: A deep visual-semantic embedding model. *Advances in neural information processing systems* 26 (2013).
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [6] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33 (2020), 6840–6851.
- [7] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685* (2021).
- [8] Yuqi Huo, Manli Zhang, Guangzhen Liu, Haoyu Lu, Yizhao Gao, Guoxing Yang, Jingyuan Wen, Heng Zhang, Baogui Xu, Weihao Zheng, et al. 2021. WenLan: Bridging vision and language by large-scale multi-modal pre-training. *arXiv preprint arXiv:2103.06561* (2021).
- [9] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*. PMLR, 19730–19742.
- [10] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International conference on machine learning*. PMLR, 12888–12900.
- [11] Junnan Li, Ramprasaath Selvaraju, Akhilesh Gotmare, Shafiq Joty, Caiming Xiong, and Steven Chu Hong Hoi. 2021. Align before fuse: Vision and language representation learning with momentum distillation. *Advances in neural information processing systems* 34 (2021), 9694–9705.
- [12] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V* 13. Springer, 740–755.
- [13] Zhuotao Liu, Yangxi Xiang, Jian Shi, Peng Gao, Haoyu Wang, Xusheng Xiao, Bihan Wen, Qi Li, and Yih-Chun Hu. 2021. Make web3. 0 connected. *IEEE transactions on dependable and secure computing* 19, 5 (2021), 2965–2981.
- [14] Ziyang Luo, Pu Zhao, Can Xu, Xiubo Geng, Tao Shen, Chongyang Tao, Jing Ma, Qingwen Lin, and Daxin Jiang. 2023. LexLIP: Lexicon-Bottlenecked Language-Image Pre-Training for Large-Scale Image-Text Retrieval. *arXiv:2302.02908 [cs.CV]*
- [15] M.L. Menéndez, J.A. Pardo, L. Pardo, and M.C. Pardo. 1997. The Jensen-Shannon divergence. *Journal of the Franklin Institute* 334, 2 (1997), 307–318. [https://doi.org/10.1016/S0016-0032\(96\)00063-4](https://doi.org/10.1016/S0016-0032(96)00063-4)
- [16] Stylianos Mystakidis. 2022. Metaverse. *Encyclopedia* 2, 1 (2022), 486–497.
- [17] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. (2008).
- [18] Bryan A Plummer, Liwei Wang, Chris M Cervantes, Juan C Caicedo, Julia Hockenmaier, and Svetlana Lazebnik. 2015. Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models. In *Proceedings of the IEEE international conference on computer vision*. 2641–2649.
- [19] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. *arXiv:2103.00020 [cs.CV]*
- [20] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. 2022. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems* 35 (2022), 25278–25294.
- [21] Quan Sun, Yuxin Fang, Ledell Wu, Xinlong Wang, and Yue Cao. 2023. EVA-CLIP: Improved Training Techniques for CLIP at Scale. *arXiv preprint arXiv:2303.15389* (2023).
- [22] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).
- [23] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
- [24] Chen-Pin Wang and Bogil Jo. 2013. Applications of a Kullback-Leibler divergence for comparing non-nested models. *Statistical modelling* 13, 5–6 (2013), 409–429.
- [25] Qin Wang, Ruija Li, Qi Wang, and Shiping Chen. 2021. Non-Fungible Token (NFT): Overview, Evaluation, Opportunities and Challenges. *arXiv:2105.07447 [cs.CR]*
- [26] Yuntao Wang, Zhou Su, Ning Zhang, Rui Xing, Dongxiao Liu, Tom H Luan, and Xuemin Shen. 2022. A survey on metaverse: Fundamentals, security, and privacy. *IEEE Communications Surveys & Tutorials* (2022).
- [27] Hu Xu, Saining Xie, Xiaoqing Ellen Tan, Po-Yao Huang, Russell Howes, Vasu Sharma, Shang-Wen Li, Gargi Ghosh, Luke Zettlemoyer, and Christoph Feichtenhofer. 2024. Demystifying CLIP Data. *arXiv:2309.16671 [cs.CV]* <https://arxiv.org/abs/2309.16671>
- [28] Yi Zheng, Yifan Zhao, Mengyuan Ren, He Yan, Xiangju Lu, Junhui Liu, and Jia Li. 2020. Cartoon face recognition: A benchmark dataset. In *Proceedings of the 28th ACM international conference on multimedia*. 2264–2272.
- [29] Zibin Zheng, Shaohan Xie, Hong-Ning Dai, Xiangping Chen, and Huaimin Wang. 2018. Blockchain challenges and opportunities: A survey. *International journal of web and grid services* 14, 4 (2018), 352–375.

Table 7: Details of NFT collections in the NFT1000 dataset

index	NFT_name	collected_tokens	index	NFT_name	collected_tokens	index	NFT_name	collected_tokens	index	NFT_name	collected_tokens	index	NFT_name	collected_tokens	index	NFT_name	collected_tokens
1	BoredApeYachtClub	10000	2	CRYPTOPUNKS	10000	3	MutantApeYachtClub	19482	4	Azuki	10000	5	ClownX	19485			
6	Moonbirds	10000	7	Doodles	10000	8	BoredApeKenneleclub	9597	9	Cool Cats	9965	10	Beanz	19950			
11	PudgyPenguins	8888	12	Cryptopaz	7024	13	World Of Women	10000	14	CyberKongz	5000	15	ONI Force	7777			
16	MekaVerse	8888	17	HAVE PRIME	8192	18	mifers	10000	19	projectPXN	10000	20	Karatu	5555			
21	My Little Friends	5000	22	Shiba Inu	10000	23	Shiba	10000	24	Pixelmon	9925	25	PixelBear	80000			
26	CyberKongz VX	14672	27	Kaijunkings	9999	28	Prime Planet	7979	29	Lazy Lions	10000	30	TLanders	9981			
31	The Doge Pound	10000	32	Deadellar	10000	33	World Of Women Galaxy	20789	34	ALIENFRENDS	10000	35	VOX Series 1	8889			
36	Hashmasks	16355	37	Psycheladies Anonymous Genesis	9595	38	VeeFriends Series 2	55554	39	RENGA	8898	40	CoolmansUnivers	10000			
41	Art Gobblers	9988	42	SupDucks	9916	43	Jungle Freaks	10000	44	Sneaky Vampire Syndicate	8888	45	SuperNormalVZipcy	8851			
46	Nakamigos	20000	47	Impostors Genesis	10420	48	Potatoz	9999	49	CryptoSkulls	10000	50	Moonbirds Odditi	10000			
51	RumbleKingLeague	10000	52	MURI	10000	53	Galactic Apes	9998	54	Lives of Asuna	9997	55	My Pet Hooligan	8888			
56	MurakamiFlowers	10105	57	Kiwami	10000	58	SHIBOSHIS	10000	59	Sappy	10000	60	DÉGEN TOONZ	8888			
61	Killer GF	7777	62	CryptоМories	5983	63	Crypto Bull Society	7777	64	CryptOrBatz by Ozzy Osbourne	9666	65	Quirkie	5000			
66	RoboCats	9999	67	Shiny Cats	9583	68	Chain Panthers	10000	69	MuturCats	9698	70	Rebeauties	9999			
71	LochChainMonkey	9601	72	Rekka	8814	73	Decadent ApeWives	10000	74	PixelBagu	2025	75	DeGodz	9066			
76	speckledclub	9999	77	The Humanoids	9901	78	Seven Temts	7000	79	Akutan	15000	80	Hypebears	10000			
81	Hero	5205	82	KIA	9998	83	inbetweeners	10777	84	C-01 Official Collection	8887	85	Imaginary Ones	8888			
86	ZombieClub Token	5478	87	Groupies	10000	88	Valhalla	9000	89	MOAR by Joan Cornellà	5555	90	Wizards & Dragons Game	45519			
91	the littles NFT	10000	92	The Heart Project	9931	93	CryptohDads	10000	94	Chimpers	5555	95	Crypto Chicks	9970			
96	VOX Series 2	8473	97	WonderPals	10000	98	LilPudgy	21243	99	A KILL called BEAST	9631	100	Akuma	5553			
101	G'EVOLs	9886	102	Tasty Bones	5049	103	Animetas	10101	104	ALACADABRAZ	9666	105	KILLABEARs	3333			
106	loomloomcraft	12345	107	Metasaurus	10000	108	Douci Darrels	10000	109	Slotie	9953	110	Party Ape Billionaire Club	5160			
111	WinterBears	10000	112	SlimHoods	4999	113	Shinsuke	1114	114	EveralDuo	7708	115	WoodMollers	5000			
116	Elves	8888	117	Apocalypic Apes	8888	118	Yakuza Club	2544	119	GalaxyFight	9994	120	BLIZZTERVERSE	8279			
121	Swampwerse	9599	122	The Doge Project	10054	123	Shiva Social Club	9998	124	Acme's Rose	10349	125	Wicked Apes	8279			
126	Cryptoh Cowen	9768	127	Anonymys	10000	128	CatBlueGenesis	9999	129	TEST NFT	10000	130	Dipped	8888			
131	Habbo Avatars	11609	132	Starchatters	9998	133	TheWickedCraniums	10762	134	Kanpal Pandas	8959	135	FoxFan	10000			
136	Gutter Dogs	2955	137	CryptonGoonz	6968	138	BASTARD GAN PUNKS V2	11303	139	HeadDAO	5552	140	Vogu	7777			
141	Boko	7777	142	Gray Boys	10000	143	Cod Monkes	10000	144	Little Lemon Friends	9999	145	Genuine Undead	9999			
146	The Other Side	8887	147	Los Muertos	10000	148	Ethlizards	9050	149	Sipher INU	10000	150	Fishy Fam	9999			
151	FOMO MOFOS	8008	152	GhostlyGhosts	6809	153	Spank Punks	10000	154	DystoPunks V2	2076	155	RareApeYachtClub	10000			
156	The Diggies	10000	157	Divine Anarchy	9929	158	Loser Club	10000	159	Fang Gang	8888	160	FrankenPunks	10000			
161	Holoset	8888	162	Weirdo Hogg Gang	5555	163	Monster Ap Club	6133	164	HOWLERZ	5000	165	Yakuza Cats Society	8929			
166	Croco Gang	8888	167	SoulzGeavers	8888	168	Wingard Ringers	10000	169	Hero Galaxies	5554	170	CreatureToadz	8888			
171	Supradive Secret Society	11110	172	Women and Weapons	10000	173	Lonely Aliens Space Club	10001	174	Kule	8260	175	Angry Ape Army	3333			
176	APE DAO REMIX!	5528	177	Foulads	4444	178	Genesis	9200	179	Alpha Kongz Club	8887	180	Sleens	10001			
181	0xApes	10146	182	ThingsDontOfficial	10000	183	Kitty Crypto Gang	7854	184	Guitar Birds	2955	185	Dapper Dinos	9997			
186	Cosmic Labs	9000	187	SIASIAN	8888	188	The Surreals	10000	189	Grandpa Ape Country Club	5000	190	Prime Kong Plane	9797			
191	CakedApez	8888	192	Crypto Cannabis Club	10000	193	Gen-F	10000	194	Pixelated Llama	3999	195	ShitBeast	10000			
196	FlowerGirls	10000	197	uwucrew	9670	198	Habibz	4900	199	Deserted Feels	11111	200	Fat Ape Club	9999			
201	Undead Pastel Club	9999	202	Mindblows	6968	203	DeadHeads	10000	204	Moornummers	9257	205	MetaBallionaire	7778			
206	Anata NFT	1993	207	Meta Eagle Club	12000	208	Mocaverse	8888	209	Hero Galaxies	5554	210	CreatureToadz	8888			
211	Doge Pound Puppies	7241	212	ZooFrenzToken	6666	213	AllToonbirds	10000	214	Box Mummy Waking Up	8888	215	MONICLES	7464			
216	Monsters	7096	217	Monsters	9999	218	Dragon's Nightmare	10000	219	Lonely Duck	10000	220	Godjira Generation 2	3333			
221	COROAKZ	6669	222	Monsters Club	8000	223	Toxic Skulls Club	9995	224	Rear Desire	6416	225	Moto-Legends	12345			
226	BullsOnTheBlock	10000	227	Goons	9697	228	BeaxR	3695	229	Onimorphs	8167	230	Alpha Girl Club	9860			
231	Claylings	4039	232	NotOkayBearz	10000	233	CryptoZuks	9991	234	Purnelopes Country Club	10000	235	Koda	3348			
236	Kumo x World	6651	237	StoneApe Club	6666	238	T THUG	7777	239	Pop Art Cats	10000	240	GenesisApostle	7363			
241	DeathApe Squad	262	242	BEANS - Dumb Ways to Die	10000	243	Metaverse	10000	244	Goofy Oversized Optics People	9999	245	Weirdo	10000			
246	Fatales	10000	247	Dragon's Claws	19549	248	Pixelmon	10000	249	Asolitiles	10000	250	Rogue Society Bots	15777			
251	HalloweenBears	9979	252	SoulZ Monogatari	7776	253	ToyBoogers	10000	254	Space Poggers	12000	255	Timeless	9411			
256	Beetings	10000	257	DuskBreakers	10000	258	Bamboozlers	9996	259	Suds NFT Heroes	5998	260	CryptohHoots Steampunk Parliament	2491			
261	CoolDudes	8888	262	BEANS - Dumb Ways to Die	10000	263	Party Penguins	10000	264	Society of Derivative Apes	9998	265	Knights of Degen	8874			
266	Outer Cool World	3060	267	Dragon's Claws	19549	268	Pixelmon	10000	269	Pixel-Apex Meta	5555	270	Pixel-Apex	8888			
271	Ghost Boyz	6666	272	Noobles	5555	273	Mutant Shiba Club	10000	274	OogaVerse	7757	275	YOLO Bunny	9999			
276	TheProjectURS	9983	277	Mutant Hounds	7011	278	Obits	7132	279	FarmerLadySquad	8888	280	Squishy Squad	8888			
281	Chill Frogs	6000	282	DystoApe	4444	283	Women Ape Yacht Club	10000	284	Meta Angels	10000	285	Avatar	2649			
286	Whiko NFT	3577	287	Woolie	9736	288	Anatomy Science Club	10000	289	Long Lost	10000	290	Kitaro World	7777			
291	0xVampire	7213	292	Superalter Apes	4444	293	BBRC OFFICIAL - IVY BOYS	7755	294	KREEPY CLUB	9999	295	Haki	5000			
296	GhostProject	10000	297	Larva Laids	5001	298	Fluffy Polar Bears	9441	299	Squishy Runners	6617	300	Unemployables	5000			
301	NinjaSquad	6888	302	FroyoKitties	10000	303	Regulars	10000	304	Space Poggers	12000	305	PeopleInThePlace	10000			
306	TrippyToadz	3067	307	The Possessed	9449	308	Fomo Dogs	10000	309	Pixelmon	7777	310	SympathyForTheDevils	4261			
311	Pixelmon	8888	312	Satoshidates	5000	313	Pixelmon	10000	314	Pixelmon	10000	315	Pixelmon	5000			
316	meeme	6000	317	Cartoons	7777	318	Pixelmon	10000	319	Pixelmon	10000	320	Khatu Mecha	2211			
321	SmallFootNFT	8888	322	MagicalVillains	8888	323	Pixelmon	10000	324	Pixelmon	10000	325	Pixelmon	2000			
326	CosmodineOmega	8888	327	UninterestedUnicorns	6888	328	Pixelmon	10000	329	Pixelmon	10000	330	Pixelmon	3000			
331	TheAlienKoy	10000	332	ElvenPrincess	5555	333	Pixelmon	10000	334	Pixelmon	10000	335	Pixelmon	10000			
336	BadBunniesNFT	5555	337	Never Fear Truth	3850	338	BlockchainBikers	10000	339	Pixelmon	10000	340	Pixelmon	10000			
341	Weather Report	10000	342	HUXLEY Robots	10000	343	RiverMen	9996	344	Pixelmon	10000	345	DIOs Genesis	4000			
346	For the Culture	6969	347	Wafus	5066	348	Lofti Originals	5555	349	Pixelmon	10000	350	Pixelmon	9899			
347	The Moon Boyz	11110	348	Chungus	8888	349	ElectricSheep	5555	350	Pixelmon	10000	351	Pixelmon	5109			
348	Heroes	3333	349	RoaringLeaders	6575	350	ShatteredEon	10000	351	Pixelmon	10000	352	Pixelmon	10000			
353	Lunartics	10000	354	Genze	9983	355	Skullpunks Hideout	10000	356	Pixelmon	9999	357	Pixelmon	9999			
358	Metakill	3000	359	Keepers V2	10001	360	Keepers Castle	2444	361	Pixelmon	10000	362	Pixelmon	10000			
363	Metakill	3000	364	Uncool Cats	6969	365	Pixelmon	10000	366	Pixelmon	10000	367	Pixelmon	10000			

Index	NFT_name	collected_tokens	index	NFT_name	collected_tokens	index	NFT_name	collected_tokens	index	NFT_name	collected_tokens	index	NFT_name	collected_tokens	index	NFT_name	collected_tokens
711	Phoenixes	8888	712	Astroheads	8845	713	HATE EXODUS	8845	714	ChillRix	9795	715	Party Grandpa Retirement Club	6000	716	Shovelz	5555
716	Scholarz	2500	717	CyberTurtles	5555	718	Silks Genesis Avatar	7302	719	Star Wolverine	8780	720	Moonscatt	5555	721	Grandpa's Garden	5000
721	Crypto Bears	7546	722	Onigiri Pepe's	6198	723	The Council	1337	724	OctoHedz V2	7998	725	ROIRO	5000	726	Cryptopunks	10001
726	CryptoApes	6969	727	Cypher Collection	3362	728	Bibos	1111	729	Pixel Interfaces	4003	730	Lazy Ape Yacht Club	10001	731	Reindeerz	6776
731	Rain Bears	2333	732	MeemoWorld	6666	733	DizzyDragons	2717	734	Broadcasters	7777	735	SuperGeisha	555	736	Space Yetis	3333
736	Soul Bears	3333	737	CRYPTONINJA WORLD	7808	738	LostSoulsSanctuary	9999	739	Japanese Born Ape Society	6883	740	Permies	555	741	Mocha Mellers	6000
741	No Funlige Frenz	6000	742	No Funlige Frenz	1001	743	GoldSilverPirates	1125	744	KumaVerse	2010	745	Tie Die Ninjas	7777	746	Lil Brains	7778
746	750	747	CornTown	10000	748	Slumdog Billionaires	1006	749	FortuneDao	658	750	Forever Fomo Duck Squad	7638	751	OKOKU	3403	
751	nobodys	3210	752	JungleLeahapechub	5555	753	AngelDevNFT	1006	754	Avius Animae	9991	755	XDX34D	6420	756	Woolly Poos	9993
756	760	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	
769	Other Duck Yacht	5555	770	Easy Demons	6666	771	MindFluff	1111	772	Shredz NFT	922	773	Survived Death	9992	774	Older Duck Yacht	6000
774	CocksDoodles	4444	775	PixelBeasts	9998	776	STRAWBERRYWTWTF	9954	777	zombieknot.wtf	7000	778	J Pierce & Friends	4000	779	Bad Bears	5555
779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	
795	DenNekkaDen Spirit Key Avatars	7743	796	LEGION	7777	797	Apocalyptic Queens	8887	798	STARCATS	1512	799	Crocodiles	8888	800	Basic Bored Apes Club	10000
800	788	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	
815	SeKings	3201	816	Character	1000	817	Bunkies	1001	818	Shark Boy Fight Club	8883	819	Doodie Dogs	10000	820	Creepz by OVERLORD	8937
820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	
836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	
852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	
868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	
884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	890	
890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	
906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	
922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	
938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	
954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	
970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	
986	987	988	989	990	991	992	993	994	995	996	997	998	999	990	991	992	
992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	
998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	
994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	
990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	
996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	
992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	
998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	
994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	
990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	
996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	
992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	
998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	
994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	
990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	
996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	
992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	
998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	
994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	
990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	
996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	
992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	
998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	
994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	
990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	
996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	
992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	
998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	
994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	
990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	
996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	
992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	
998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	
994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	
990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	
996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	
992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	
998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	
994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	
990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	
996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	
992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	
998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	
994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	
990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	
996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	
992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	
998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	
994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	
990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	
996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	
992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	
998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	
994	995	996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	
990	991	992	993	994	995	996	997	998	999	990	991	992	993	994	995	996	
996	997	998	999	990	991	992	993	994	995	996	997	998	999	990	991	992	
992	993	994	995	996	99												