

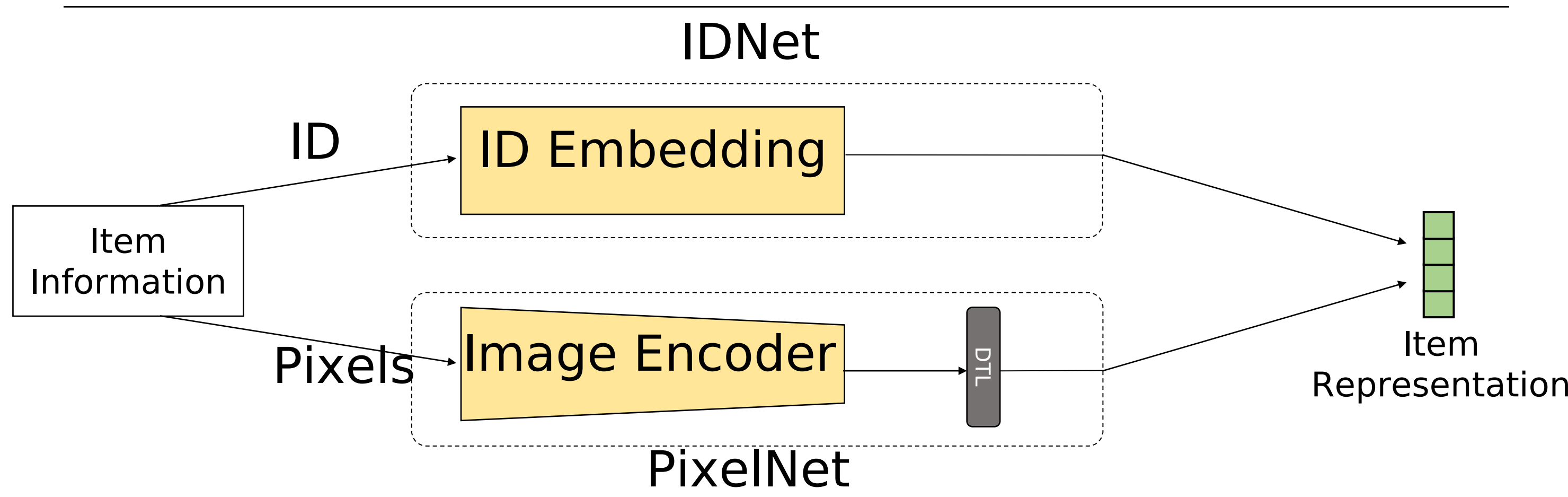
An Image Dataset for Benchmarking Recommender Systems with Raw Pixels

Yu Cheng^{1,2}, Yunzhu Pan², Jiaqi Zhang², Yongxin Ni², Aixin Sun³, and Fajie Yuan²

¹ Zhejiang University ² Westlake University ³ Nanyang Technological University



Motivation



Encouraged by the success of LLMs
ID-based Recsys → Modality-based Recsys
Targeting at foundation models in Recsys

Key weaknesses of existing visual recommender dataset:
(1) Mismatch in tasks and vocabs (2) Hindering technological advancement in CV (3) Limited scale .

	Flickr	Behance	KuaiRec	Pinterest	WikiMedia	Amazon	GEST	HM	PixelRec
Raw Image		×		✓	✓		✓		✓
Large Scale		×		×	×		✓		✓
Scene		MM		MM		e-commerce			MM

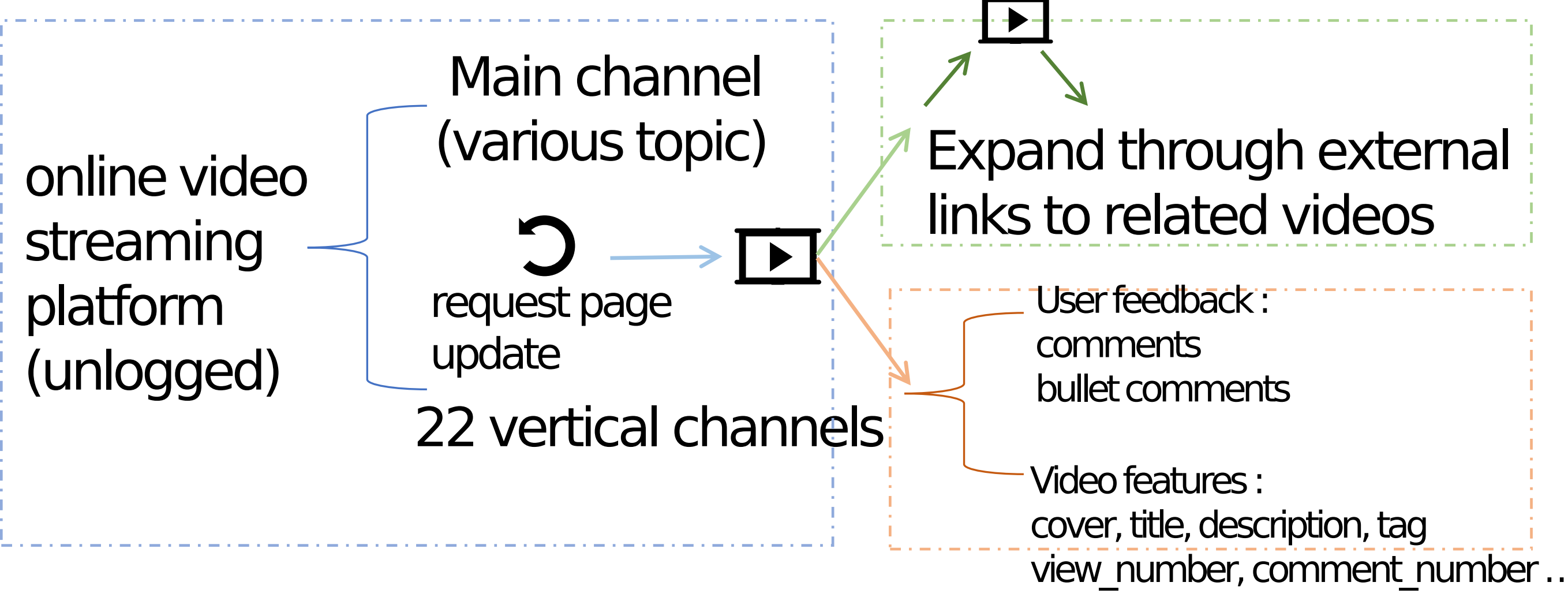
PixelRec Dataset

PixelRec : A large scale recommendation dataset collected from short video platform, supporting :

- End-to-End training for visual recommendation models
- Effective and precise image-based recommendation
- Short-video/ multimodal recommendation
- Pre-training resource for foundation vision recommendation models
- Studying preference models founded solely on images
- Facilitating Bridging of RS and CV domains

- **30 million** Users
- **400,000** high quality images
- **200 million** Interactions
- Abundant features for item

Dataset Construction



Contribution

PixelRec Benchmark on Recommender Backbone

ItemEnc	Metrics	Non-Sequential Recommender			Sequential Recommender					
		MF	FM	DSSM	SRGNN	GRU4Rec	BERT4Rec	NextItNet	SASRec	LightSANs
ID	Recall@10	1.013	1.357	1.401	1.597	1.833	1.972	2.187	2.500	2.578
	NDCG@10	0.490	0.679	0.701	0.808	0.937	0.994	1.153	1.350	1.384
RN50	Recall@10	0.357	1.024	0.960	2.224	2.294	2.391	2.140	2.633	2.417
	NDCG@10	0.169	0.501	0.475	1.132	1.138	1.199	1.073	1.321	1.226
ViT	Recall@10	0.472	1.124	1.242	2.152	2.102	2.450	2.215	2.583	2.461
	NDCG@10	0.229	0.543	0.617	1.065	1.031	1.230	1.106	1.292	1.224

- Nine Recommender Backbones
- Nine Image Encoders
- Exhaustive search on hyper-parameters

Exploratory Results

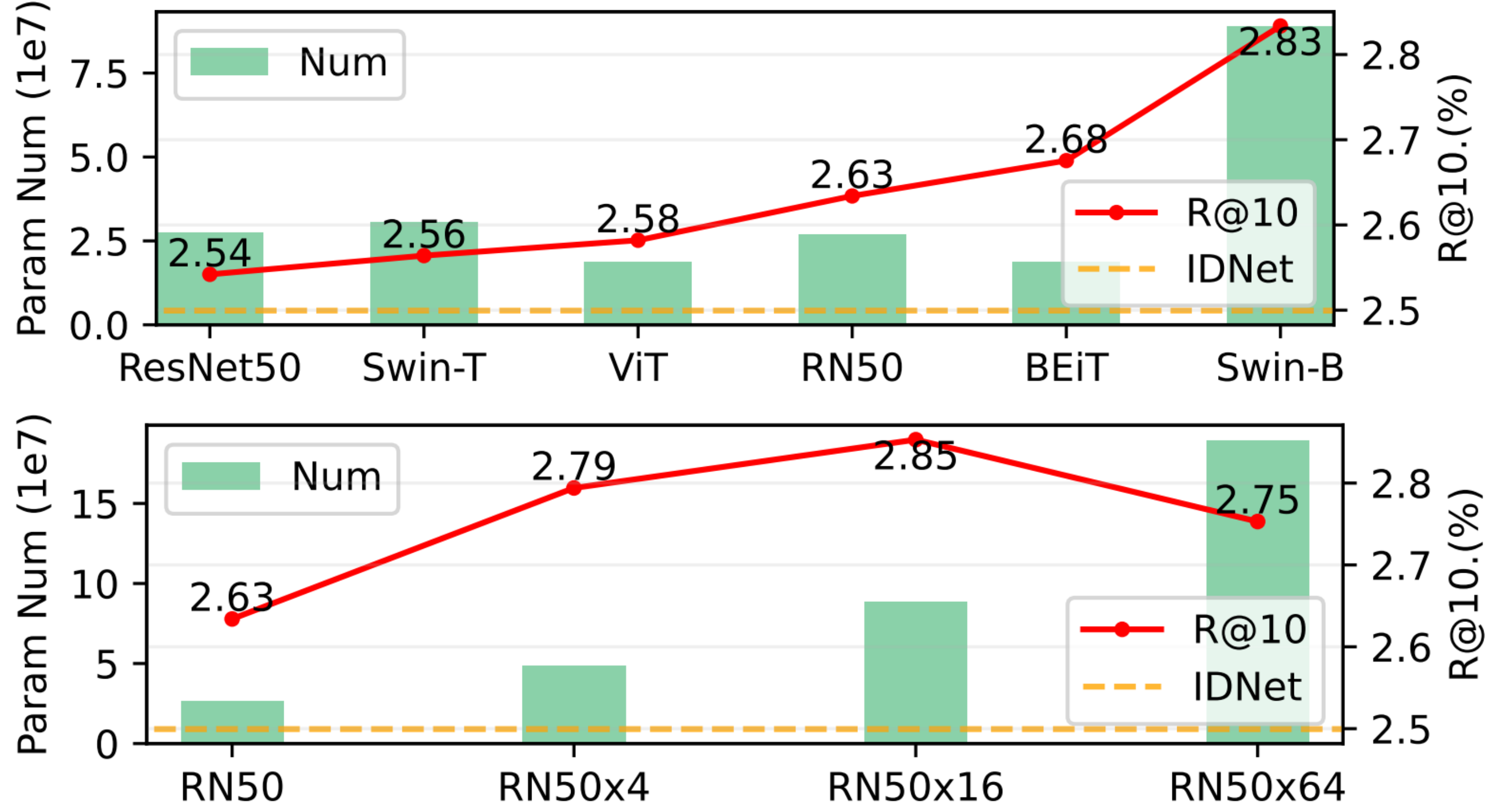
• Adopting sequential recommender backbone and end2end training strategy, PixelNet perform satisfactorily in regular recommendation setting

• The performance of PixelNet may be significantly influenced by the specific recommendation backbone network and training approach used

Future works

- Reducing computation consumption of end2end training
- Effective hyper-parameter tuning of PixelNet
- Building foundation vision recommender models

PixelRec Benchmark on Image Encoder



Dataset Overview

Title:
[4K Healing] Anime scenes that exist in real life...

Description:
Extended version of this video, "Enter the World Inside the Wallpaper"...

Release time: 2021-7-17

Video tag: trip shoot

view number: 3951 K **comment number:** 5732 **coin number:** 136 K

favorite number: 232 K **share number:** 93 K **barrage number:** 5121