

# DH-FaceVid-1K: A Large-Scale High-Quality Dataset for Face Video Generation

Donglin Di<sup>1</sup>, He Feng<sup>2</sup>, Wenzhang Sun<sup>1</sup>, Yongjia Ma<sup>1</sup>, Hao Li<sup>1</sup>, Wei Chen<sup>1</sup>, Lei Fan<sup>3</sup>, Tonghua Su<sup>2</sup>, Xun Yang<sup>4\*</sup>

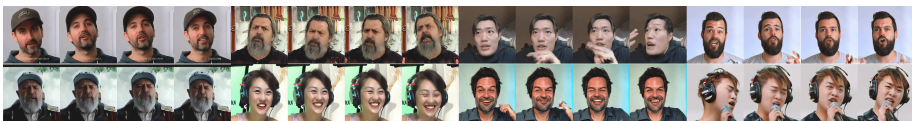
<sup>1</sup>Li Auto, <sup>2</sup>HIT, <sup>3</sup>UNSW, <sup>4</sup>USTC

## DH-FaceVid-1K Dataset



**Ethnic Diversity:** Features Asian majority (>80%) while keeping ethnic diversity. **High-Resolution:** All videos are  $\geq 512 \times 512$ , with 25% exceeding 1080p. **Multi-modal Annotations:** Each sample is paired with speaking audio and text. **Meticulously Curated:** Explicitly filtered to common flaws like hand occlusions.

Datasets	#_Samples	Resolution	Duration	Text	Audio
VoxCeleb2	1,092,008	224×224	2400h	X	✓
HDFT	368	512×512 +	15.8h	✓	✓
TalkingHead-1KH	80,000	512×512 +	160h	✓	✓
CelebV-HQ	35,666	512×512 +	68h	✓	✓
CelebV-Text	70,000	512×512 +	279h	✓	✓
DH-FaceVid-1K	270,043	512×512 +	1200h	✓	✓



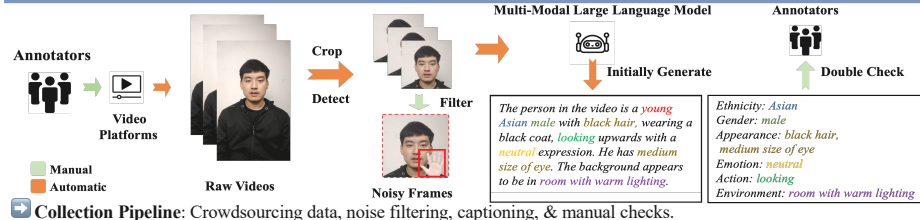
**Persistent Artifacts:** Generated results still suffer from flaws like deformed hands.

## Quantitative Results

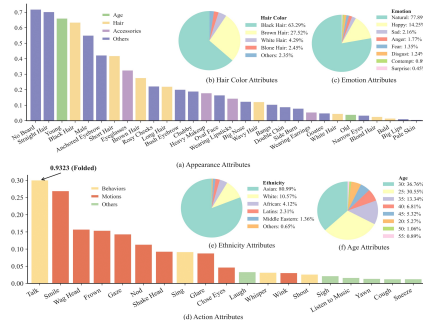
	AnimateDiff [22]			Latte [37]			OpenSora [79]			EasyAnimate [63]			CogVideoX [71]		
Datasets	FVD (↓)	FID (↓)	CLIP (↑)	FVD (↓)	FID (↓)	CLIP (↑)	FVD (↓)	FID (↓)	CLIP (↑)	FVD (↓)	FID (↓)	CLIP (↑)	FVD (↓)	FID (↓)	CLIP (↑)
HDFT	377.17	58.52	0.3447	175.01	25.89	0.3319	143.52	20.01	0.3171	137.10	19.05	0.3212	127.88	17.83	0.3237
TalkingHead-1KH	280.57	52.22	0.3368	235.06	28.44	0.3433	176.43	21.56	0.3120	166.18	19.77	0.3236	167.23	19.02	0.3223
CelebV-HQ	257.61	41.42	0.3406	169.43	25.05	0.3219	142.32	18.08	0.3374	147.62	17.45	0.3205	137.62	17.31	0.3408
CelebV-Text	218.93	35.22	0.3423	159.52	26.83	0.3372	131.06	17.44	0.3355	121.22	16.53	0.3474	129.59	15.82	0.3440
DH-FaceVid-1K	176.13	28.69	0.3557	140.01	24.95	0.3684	117.88	15.89	0.3505	113.27	13.91	0.3340	98.01	11.73	0.3801

**Improved Performance:** Attributed to larger data scale and manual noise filtering.

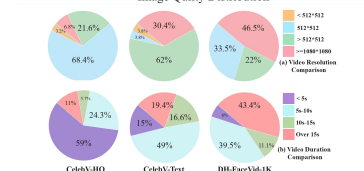
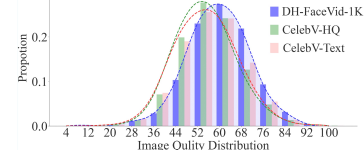
## Collection and Attributes



**Collection Pipeline:** Crowdsourcing data, noise filtering, captioning, & manual checks.



**Attribute Distribution:** Covering a wide range of ethnicities, ages, and emotions.



**Visual Quality:** Featuring superior fidelity and a high-resolution focused distribution.

## Access Dataset



### Dataset Access Workflow:

- Scan the QR code to access and agree user agreement.
- Submit the form with your intended use and institutional email.



Await our manual review of your application.

Receive download links via your institutional email upon approval.

Contact us: [fenghe021209@gmail.com](mailto:fenghe021209@gmail.com)