

Planned intervention: On Wednesday, July 16th 05:00 UTC Zenodo will be unavailable for 15-30 minutes to perform a storage cluster upgrade.



Published March 24, 2025 | Version 1.0

Dataset Open

Mediapipe based Preprocessed VGGFace2 Dataset

Shah, Syed Taimoor Hussain (Researcher)¹ ; Shah, Syed Adil Hussain (Researcher)^{2, 1} ;
Zamir, Ammara (Researcher)³ ; Qayyum, Kainat (Researcher)⁴ ;
Shah, Syed Baqir Hussain (Researcher)⁵ ; Fatima, Syeda Maryam (Researcher)⁶ ;
Deriu, Marco Agostino (Supervisor)⁷

Show affiliations

VGGFace2 Dataset and Face Mesh Preprocessing

Introduction

The VGGFace2 dataset is a large-scale face recognition dataset containing over 3.31 million images of 9,131 identities, with an average of 362 images per identity. The dataset is designed to include extensive variations in pose, age, illumination, ethnicity, and profession, making it one of the most diverse and challenging face recognition datasets available. For more details, please refer to the original publication:

VGGFace2: A dataset for recognizing faces across pose and age - DOI: 10.48550/arXiv.1710.08092

Preprocessing Using MediaPipe 3D Face Mesh

On this dataset, we applied the MediaPipe-based 3D face mesh algorithm to accurately detect faces while removing all background elements, including hair. Our preprocessing strictly retained facial landmarks, ensuring that only the essential facial features were preserved. This approach significantly enhanced the accuracy and generalization of our model, as the model was trained exclusively on landmark-based facial data.

Training and Performance

The preprocessed data was utilized to train Xception model, which resulted in remarkably accurate outcomes due to the strictly landmark-based facial representation. The model demonstrated robust performance including explainable-AI, proving that eliminating unnecessary background elements contributed positively to its efficiency and reliability.

Citation

If you use this dataset or the preprocessed version in your work, please cite both of the following:

VGGFace2 Dataset:

```
@article{Cao2018VGGFace2,  
  title={VGGFace2: A dataset for recognizing faces across pose and age},  
  author={Cao, Qiong and Shen, Li and Xie, Weidi and Parkhi, Omkar M and Zisserman, Andrew},  
  journal={arXiv preprint arXiv:1710.08092},
```

```
year={2018}  
}
```

DOI: [10.48550/arXiv.1710.08092](https://doi.org/10.48550/arXiv.1710.08092)

Preprocessed Dataset using MediaPipe:@dataset{Shah2025_MediaPipe_FaceMesh,

title={MediaPipe-based 3D Face Mesh Preprocessed VGGFace2 Dataset},

author={Shah, Syed Taimoor Hussain and Shah, Syed Adil Hussain and Zamir, Ammara and Qayyum, Kainat and Shah, Syed Baqir Hussain and Fatima, Syeda Maryam and Deriu, Marco Agostino},

year={2025},

doi={10.5281/zenodo.15078557}

```
}  
DOI: [10.5281/zenodo.15078557](https://doi.org/10.5281/zenodo.15078557)
```

Contact

For any questions or further details, please feel free to contact us.

Syed Taimoor Hussain Shah

PolitoBIOMed Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Turin, Italy

Email: taimoor.shah@polito.it

ORCID: 0000-0002-6010-6777

Files

VGGFace2.zip
















 VGGFace2.zip



! The previewer is not showing all the files.

■ VGGFace2

■ A._A._Gill_Male

 0001_01.jpg	22.7 kB
 0003_01.jpg	24.2 kB
 0004_02.jpg	24.1 kB
 0005_01.jpg	24.2 kB
 0008_01.jpg	23.2 kB
 0009_02.jpg	24.2 kB
 0014_01.jpg	23.3 kB
 0015_01.jpg	23.6 kB
 0017_01.jpg	23.5 kB
 0028_01.jpg	25.0 kB
 0029_01.jpg	22.8 kB
 0033_01.jpg	21.2 kB
 0043_01.jpg	24.9 kB

Files (11.5 GB) >

VGGFace2.zip

md5:6b18eec410a88aa5d0b142b820453b96 

11.5 GB

 Preview

 Download

exception_training_history_accuracy.jpg

md5:1151b5f9de7ee9cc45d94a63aa8c76e0 

329.9 kB

 Preview

 Download

[xception_training_history_loss.jpg](#)md5:9c6ed31d493a6e69073a5a0fac60fbf1 

301.5 kB

 Preview Download

Additional details

Funding

European Commission

PARENT – PremAtuRe nEwborn motor and cogNitive impairmenTs: Early diagnosis

956394

[Citations](#) **Show only:**

Search for citation ...

Search

 Literature (0) Dataset (0) Software (0) Unknown (0) Citations To This Version*No citations found***226**
 **VIEWS****100**
 **DOWNLOADS**[▶ Show more details](#)

Versions

Version 1.0

Mar 24, 2025

10.5281/zenodo.15078557

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.15078556](https://doi.org/10.5281/zenodo.15078556). This DOI represents all versions, and will always resolve to the latest one. [Read more](#).

External resources

Indexed in



Communities



PARENT H2020

Details

DOI

DOI [10.5281/zenodo.15078557](https://doi.org/10.5281/zenodo.15078557)

Resource type

Dataset

Publisher

Zenodo

Rights

License



Creative Commons Attribution 4.0 International

Citation

Shah, S. T. H., Shah, S. A. H., Zamir, A., Qayyum, K., Shah, S. B. H., Fatima, S. M., & Deriu, M. A. (2025). Mediapipe based Preprocessed VGGFace2 Dataset (1.0) [Data set]. Zenodo.

<https://doi.org/10.5281/zenodo.15078557>

Style

APA



Export

JSON



Export

Technical metadata

Created March 24, 2025

Modified March 24, 2025



Jump up

About

Blog

Help

Developer
s

Contribute

Funded by

About

Blog

FAQ

REST API

 GitHub

Policies

Docs

OAI-PMH

 Donate



Infrastructure

Guides

Principles

Support

Projects

Roadmap

Contact

Powered by CERN Data Centre & InvenioRDM

[Status](#) [Privacy policy](#) [Cookie policy](#) [Terms of Use](#)

[Support](#)