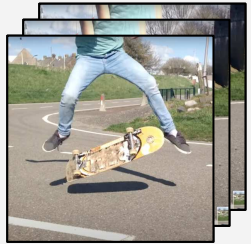
This is an image of a flamingo.

Image-Text Pairs dataset

[N=1, T=1, H, W, C]



A kid doing a kickflip.

Video-Text Pairs dataset

[N=1, T>1, H, W, C]

Welcome to my website!

This is a picture of my dog.

This is a picture of my cat.

Multi-Modal Massive Web (M3W) dataset

[N>1, T=1, H, W, C]