Single-frame Training Process

Image (RGB)

↓

Dataset
Data Augmentation
out: [C, H, W]

↓

Dataloader
divides into batches
out: [B, C, H, W]

↓

Model
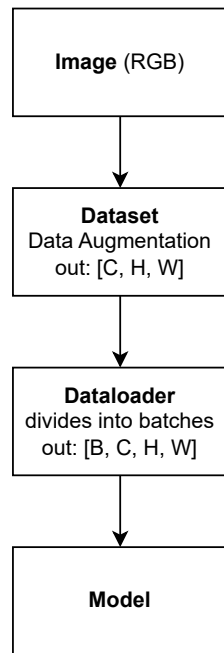
Temporal Training Process

Video
(S Images (RGB))

↓

Dataset
Data Augmentation
out: [S, C, H, W]

↓

Dataloader
divides into batches
out: [B, S, C, H, W]

←→

Sampler
indexing for
sequences

↓

Temporal Model