Fine-tuning with limited voltage-tactile pairs

Paired voltage measurement **Paired tactile map**

E2IM

Masked *

$\mathbf{v}_p$

P.E. $\longrightarrow \oplus$

$\mathbf{z}_0$

Tokenization

$\mathbf{z}_1$

Pretrained Voltage Encoder ❄

$\mathcal{E}_{vol}$

$\mathbf{z}_v$

Linear Projection

$\mathbf{z}_t$

Pretrained Tactile Decoder 🔥

$\mathcal{D}_{tact}$

S *

$\mathcal{L}_{tact}$

Tactile map

Pretraining with E2IM

Patch + Positional Embedding (P.E.)

$\mathcal{E}_{vol}$ $\mathcal{D}_{vol}$

Transformer ... Transformer

12

ViT-based Encoder

Transformer Transformer

2

ViT-based Decoder

Positional Embedding

$\mathbf{v}$ $\hat{\mathbf{v}}$

$\mathcal{L}_{vol}$

Pretraining with tactile map

$\mathcal{E}_{tact}$ $\mathcal{D}_{tact}$

ResNet-based Encoder

Self attention   ResNet-based Decoder

$\mathbf{t}$ $\hat{\mathbf{t}}$

$\mathcal{L}_{tact}$