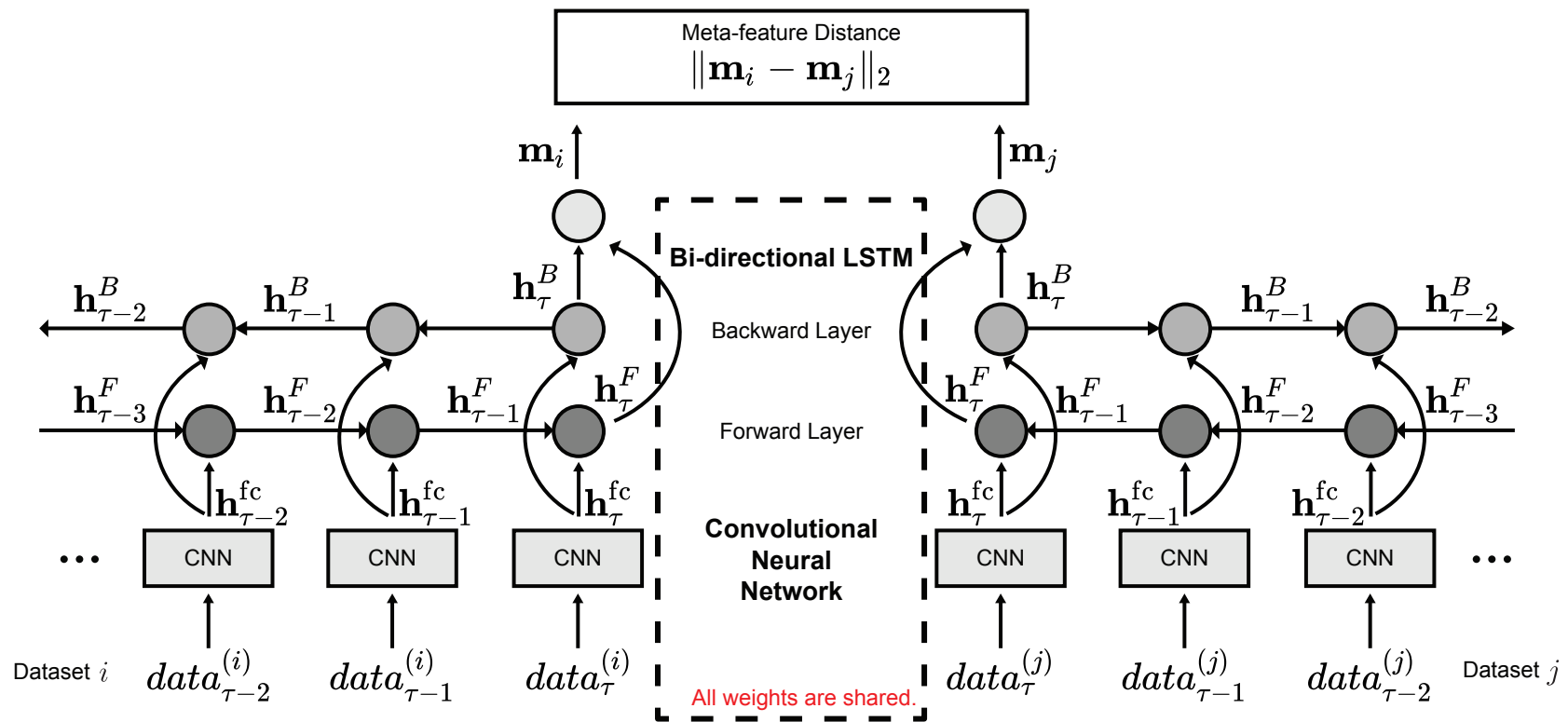# Learning to Transfer Initializations for Bayesian Hyperparameter Optimization
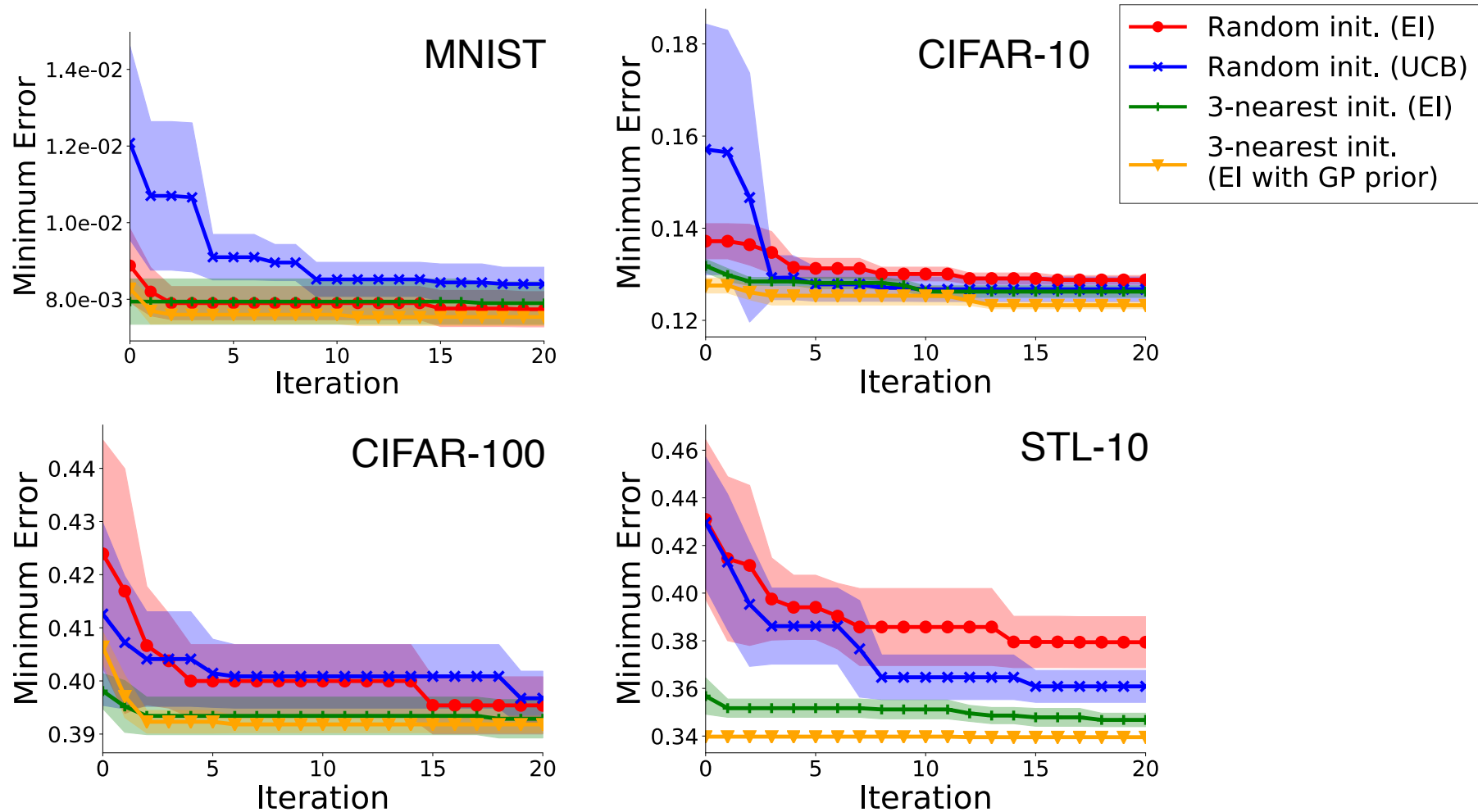
Jungtaek Kim        POSTECH

**Saehoon Kim        Altrics / POSTECH (presenter)**

Seungjin Choi       POSTECH

AItrics

# 01 Proposed Method: Meta-Feature Learning with Siamese Architecture

Meta-feature Distance
$$\|\mathbf{m}_i - \mathbf{m}_j\|_2$$

$\mathbf{m}_i$       $\mathbf{m}_j$

**Bi-directional LSTM**

$\mathbf{h}^B_{\tau-2}$   $\mathbf{h}^B_{\tau-1}$   $\mathbf{h}^B_\tau$    Backward Layer    $\mathbf{h}^B_\tau$   $\mathbf{h}^B_{\tau-1}$   $\mathbf{h}^B_{\tau-2}$

$\mathbf{h}^F_{\tau-3}$   $\mathbf{h}^F_{\tau-2}$   $\mathbf{h}^F_{\tau-1}$   $\mathbf{h}^F_\tau$    Forward Layer    $\mathbf{h}^F_\tau$   $\mathbf{h}^F_{\tau-1}$   $\mathbf{h}^F_{\tau-2}$   $\mathbf{h}^F_{\tau-3}$

$\mathbf{h}^{\mathrm{fc}}_{\tau-2}$   $\mathbf{h}^{\mathrm{fc}}_{\tau-1}$   $\mathbf{h}^{\mathrm{fc}}_\tau$    **Convolutional Neural Network**    $\mathbf{h}^{\mathrm{fc}}_\tau$   $\mathbf{h}^{\mathrm{fc}}_{\tau-1}$   $\mathbf{h}^{\mathrm{fc}}_{\tau-2}$

$\cdots$   CNN   CNN   CNN     CNN   CNN   CNN   $\cdots$

Dataset $i$   $data^{(i)}_{\tau-2}$   $data^{(i)}_{\tau-1}$   $data^{(i)}_\tau$    All weights are shared.    $data^{(j)}_\tau$   $data^{(j)}_{\tau-1}$   $data^{(j)}_{\tau-2}$   Dataset $j$

$$\mathcal{L}(\mathcal{D}^{(i)}, \mathcal{D}^{(j)}) = \left[ d_{\mathrm{target}}(\mathcal{D}^{(i)}, \mathcal{D}^{(j)}) - \|\mathbf{m}_i - \mathbf{m}_j\|_2 \right]^2$$

AItrics

# 02 Experimental Results

Thank you for listening!
Please stop by our poster.