

Supporting Information

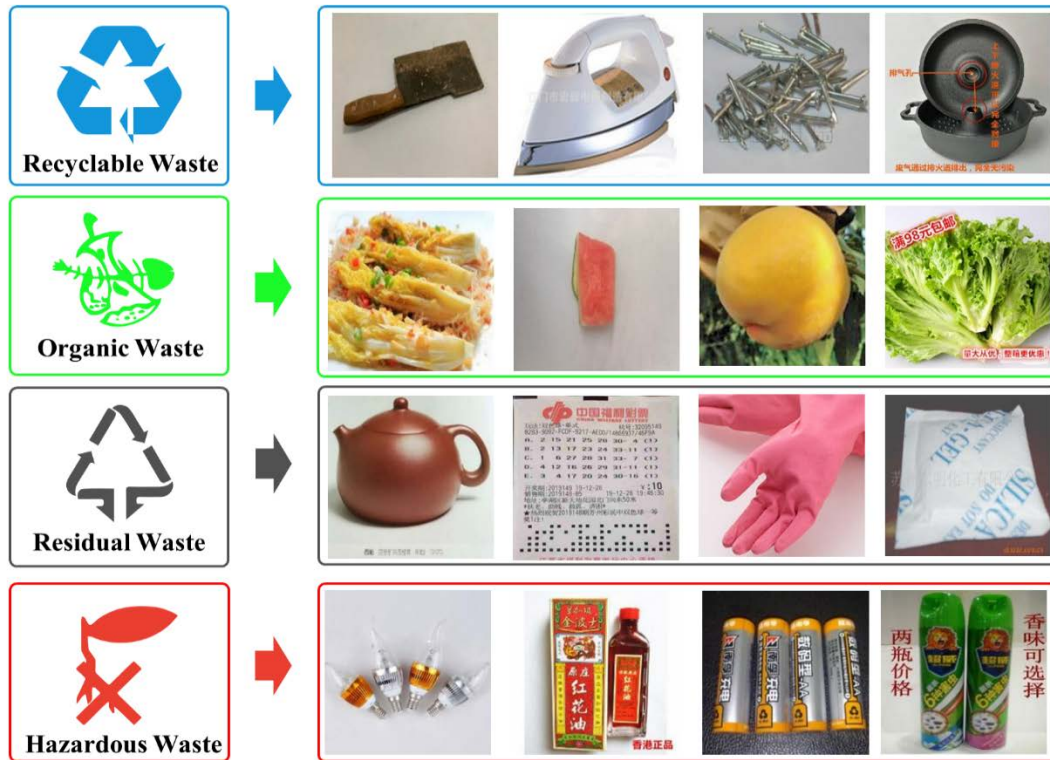


Fig. S1 Example of recyclable waste, organic waste, residual waste, and hazardous waste.

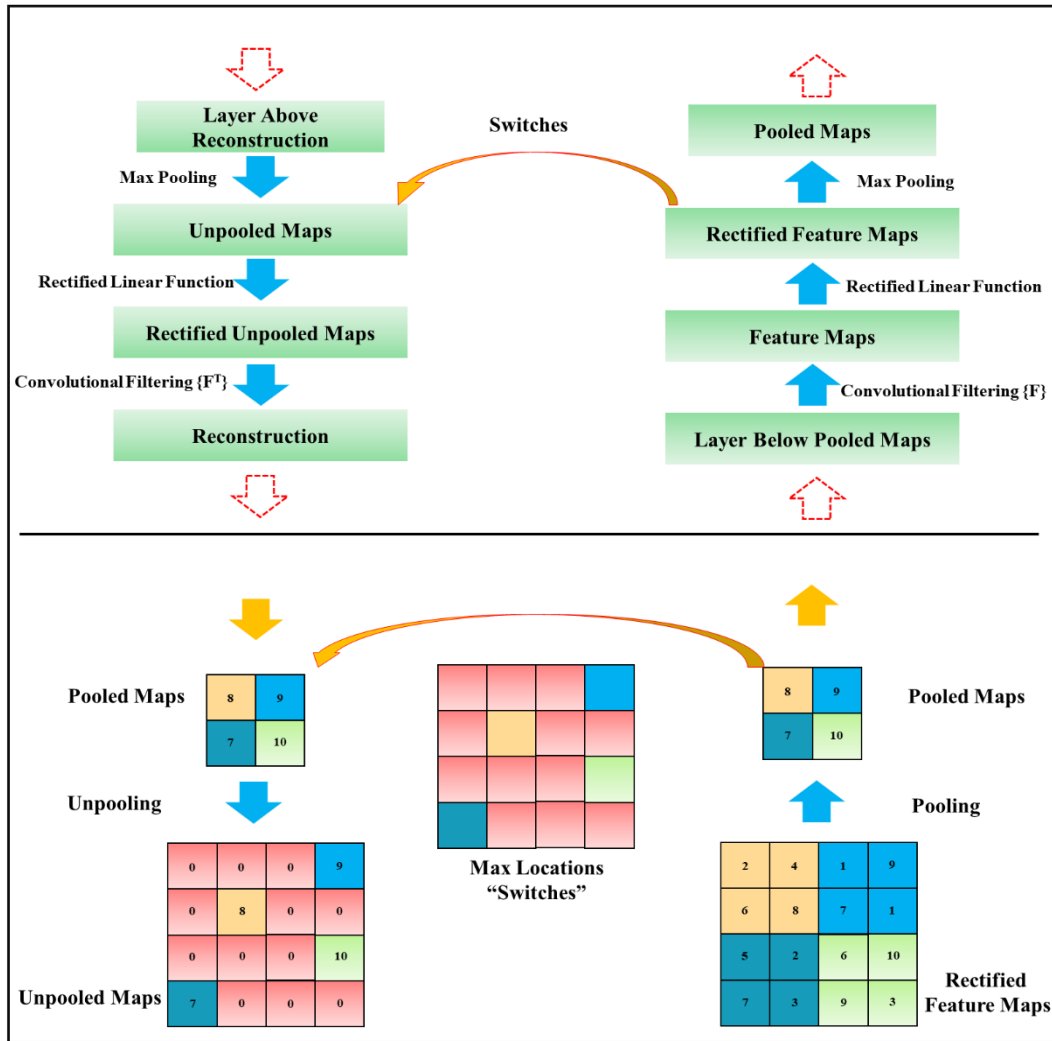


Fig. S2 Mechanism of realizing visualization with a deconvolution network.

Top: A deconvolutional layer (left) attached to a convnet layer (right). Bottom: an illustration of the un-pooling operation in the deconvolutional layer, using switches that record the location of the local max in each pooling region during pooling in the convolutional layer.

Mechanism of cyclical learning rate

$$\text{local cycle} = \text{math} \cdot \text{floor}(1 + \text{EC} / (2 \times \text{Stepsize})) \quad (\text{S-1})$$

$$\text{local } x = \text{math} \cdot \text{abs}(\text{EC} / (\text{Stepsize}) - 2 \times \text{cycle} + 1) \quad (\text{S-2})$$

$$\text{local lr} = \text{opt.LR} + (\text{maxLR} - \text{opt.LR}) \times \text{math} \cdot \text{max}(0, (1 - x)) \quad (\text{S-3})$$

where EC and Stepsize represent the number of epochs of training and half the period or cycle length, respectively; lr was represented learning rate; opt.LR is the specified lower learning rate such as min_lr; in addition, maxLR is represented the maximum learning rate.

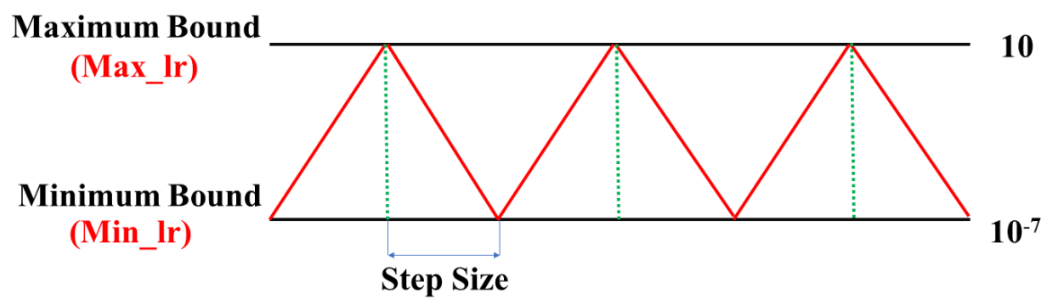


Fig. S3 Triangular rate policy. The red lines represent learning rate values changing between bounds.

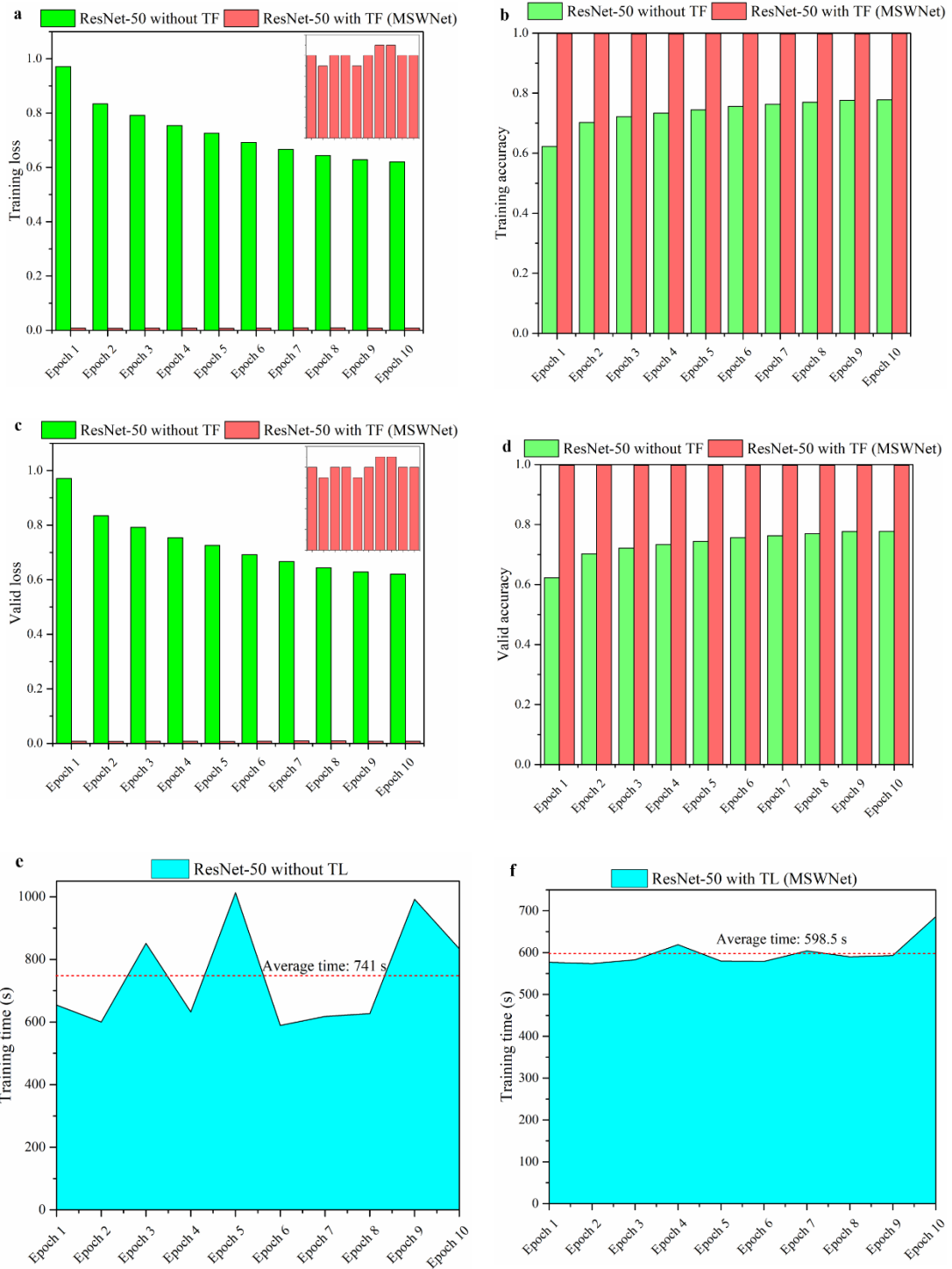


Fig. S4 Comparison of the performance of waste sorting using ResNet-50 with/without transfer learning: (a) Training loss; (b) Training accuracy; (c) Validating loss; (d) Validating accuracy; (e) Training time in ResNet-50; (f) Training time in MSWNet. Note: TL was represented to transfer learning.