
Unlearning dataset $\mathcal{D}_f$
Adversarial examples $\widetilde{\mathcal{D}}_f$
$\mathcal{D}_f$
[0.15, 0.8, 0.05]
automobile
Large importance
Small importance
More change
Less change