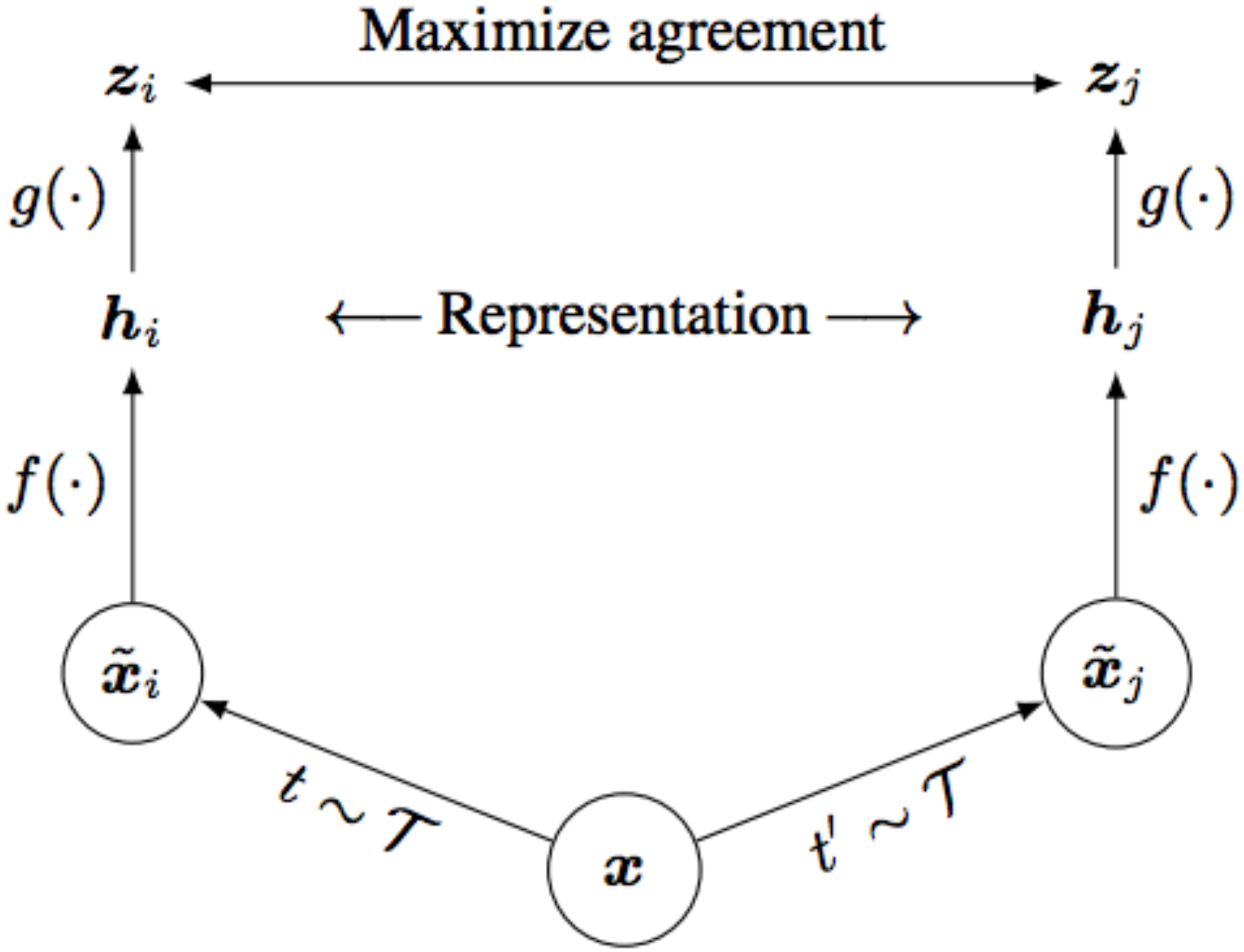
A Simple Framework for Contrastive Learning of Visual Representation

"SimCLR"

The Contrastive Learning Framework



The Contrastive Learning Framework (1)

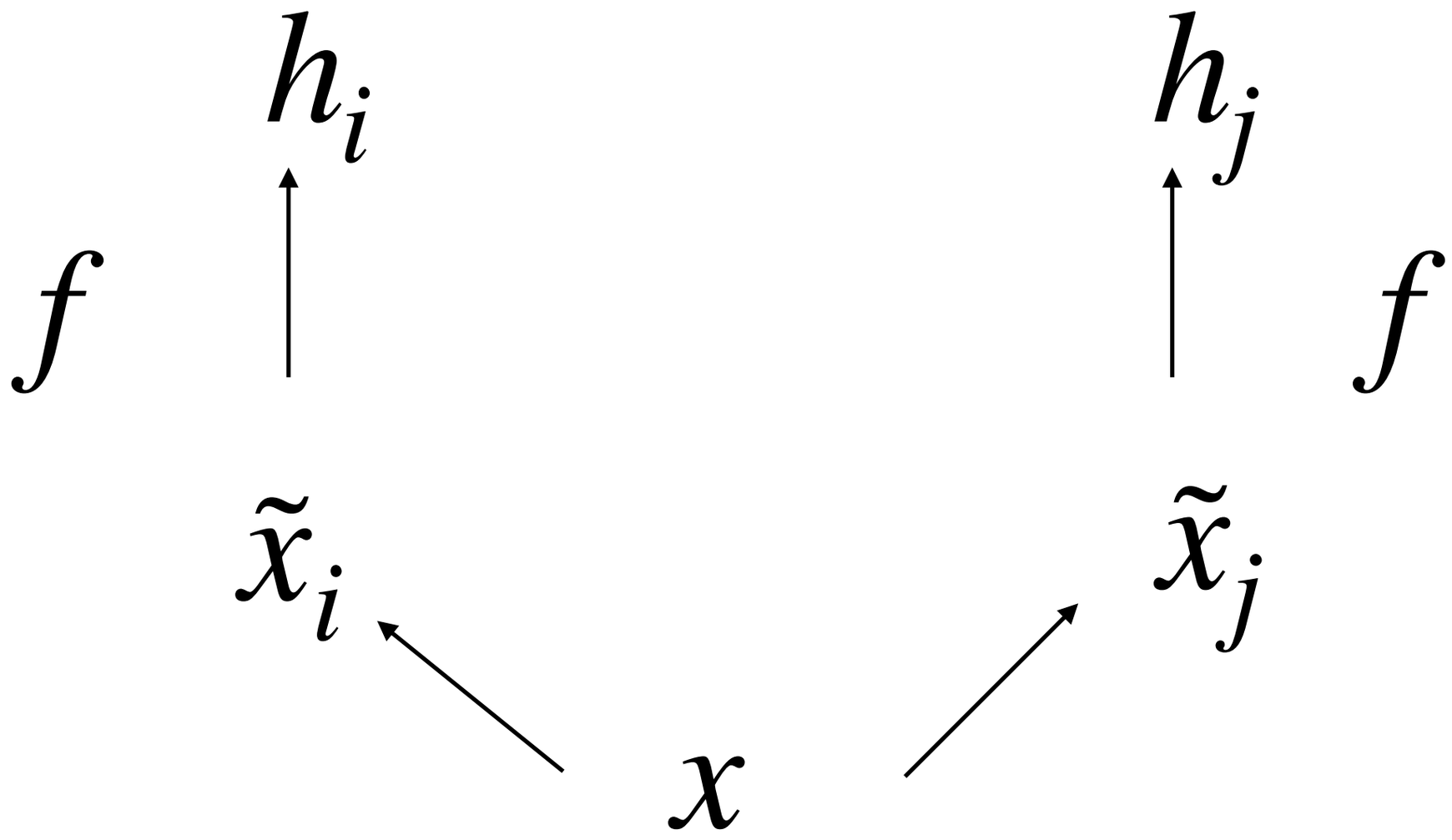
stochastic data augmentation module

$\tilde{x}_i$ $\tilde{x}_j$

x

The Contrastive Learning Framework (2)

base encoder that extracts representation vectors



The Contrastive Learning Framework (1)

small neural network projection head

z_i z_j

g g

h_i h_j

The Contrastive Learning Framework (4)

z_i **maximize agreement via contrastive loss** z_j

contrastive prediction task:

given $\{\tilde{x}_k\}$

identify $\tilde{x}_j$ **in** $\{\tilde{x}_k\}_{k \neq i}$ **for a given** $\tilde{x}_i$

NT-Xent Loss Function

$$l_{i,j} = -\log \frac{exp(sim(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} 1_{k \neq i} exp(sim(z_i, z_j)/\tau}$$

temperature parameter

indicator function

randomly sample a minibatch of N examples

the final loss is computed across all positive pairs, both (i,j) and (j,i)

SimCLR main learning algorithm

input: batch size N, structure of f, g, $\mathcal{T}$
for all sampled minibatch $\{x_k\}_{k=1}^{N}$ **do**
    for all k **do**
        draw two augmentation functions $t \sim \mathcal{T}$, $t' \sim \mathcal{T}$
        the first augmentation
        representation f
        projection g
        the second augmentation
        representation f
        projection g
    end for
end for
for all $i, j \in \{1, \ldots, 2N\}$ **do**
    calculate loss
    update networks f, g to minimize loss
end for