本文解决ood问题的思想是建模domain的特征统计量的不确定性：不把这些特征统计量当成确定值，而是假设它们服从多元高斯分布，即统计量的均值和标准差分别服从潜在的高斯分布$\mathcal{N}(\mu, \Sigma_\mu)$和$\mathcal{N}(\sigma, \Sigma_\sigma)$，且该潜在高斯分布的中心是在各统计量处，也就是说$\mu$和$\sigma$是容易得到的，取统计量即可。关键在于对$\Sigma_\mu$和$\Sigma_\sigma$的建模，这两个统计量的方差体现了数据分布的不确定性，即domain shift的方向和强度。**因此本文的思路大致可以理解为通过建模分布的分布来刻画ood数据。**据此提出了一种新的data augmentation的方法，通过对数据统计量的重参数化技巧，建模在domain shift条件下，统计特征的不确定性。然后通过得到的概率模型对数据进行重新采样，实现data augmentation，以增强网络的泛化能力。

**具体方法：**

给定$x \in R^{B\times C\times H\times W}$，$B$是batch大小，$C$是channel数（为什么是分channel而不是domain？），首先计算channel wise的均值$\mu \in R^{B\times C}$和方差$\sigma^2 \in R^{B\times C}$：$\mu(x) = \frac{1}{HM}\sum_{h=1}^{H}\sum_{w=1}^{W} x_{b,c,h,w}$ $\sigma^2(x) = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}(x_{b,c,h,w} - \mu(x))^2$

**得到的只是基于训练数据的统计量，或者可以理解为特征变化的中心。而变化的幅度，即$\Sigma_\mu$和$\Sigma_\sigma$要靠下面的方法确定：**

提出了一种非参数化的方法，即对batch内的数据算方差：$\Sigma_\mu^2 = \frac{1}{B}\sum_{b=1}^{B}(\mu(x) - E_b[\mu(x)])^2$ $\Sigma_\sigma^2 = \frac{1}{B}\sum_{b=1}^{B}(\sigma(x) - E_b[\sigma(x)])^2$

得到的$\Sigma_\mu \in R^C$和$\Sigma_\sigma \in R^C$是对各channel的均值和标准差的不确定性的估计。

之后可以用重参数化技巧建模数据均值和标准差的分布：
$\beta(x) = \mu(x) + \epsilon_\mu \Sigma_\mu(x),\ \ \epsilon_\mu \sim \mathcal{N}(0,1) \gamma(x) = \sigma(x) + \epsilon_\sigma \Sigma_\sigma(x),\ \ \epsilon_\sigma \sim \mathcal{N}(0,1)$

然后抽样得到增强后的数据$\hat{x} = \gamma(x) \times \frac{x - mu(x)}{\sigma(x)} + \beta(x)$。实际中，以p的概率执行如上的数据增强。这种操作可以灵活地加入网络的任何位置。