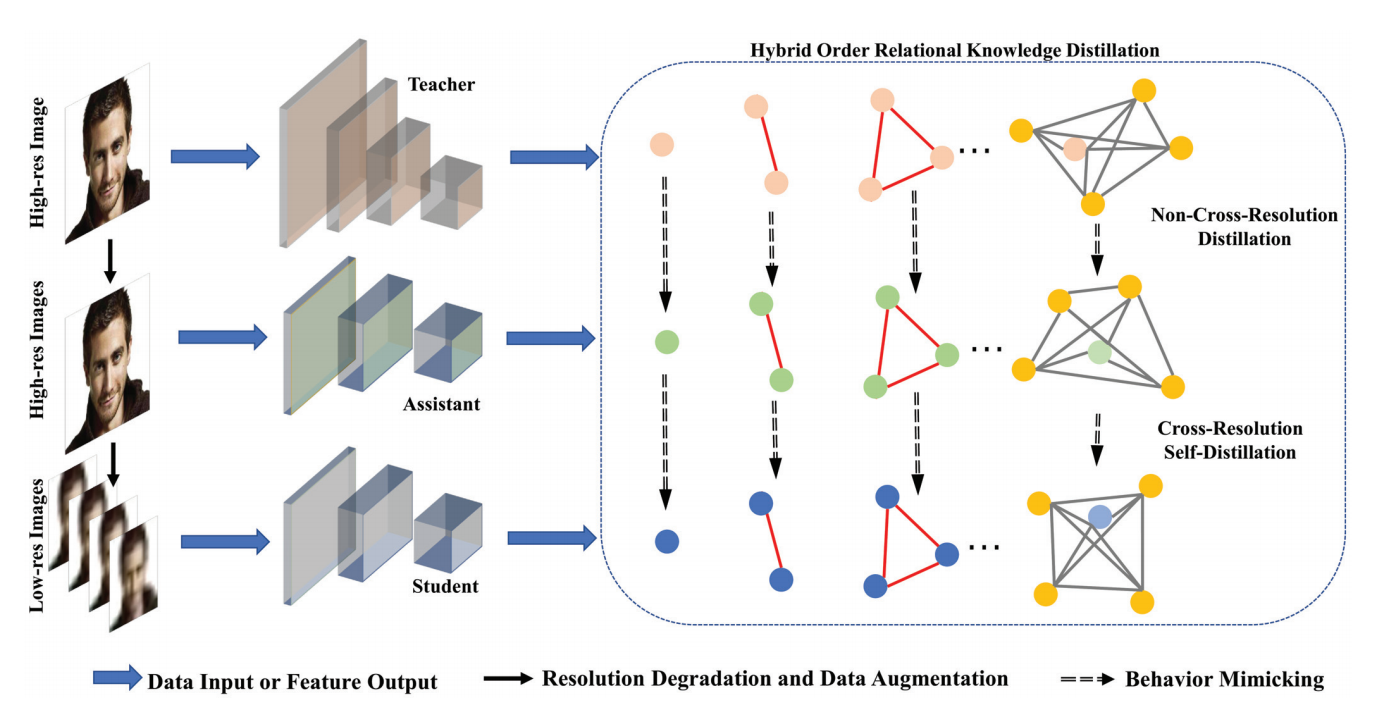
一目了然：提取混合阶数关系知识以进行跨分辨率图像识别

**Look One and More: Distilling Hybrid Order Relational Knowledge for Cross-Resolution Image Recognition**

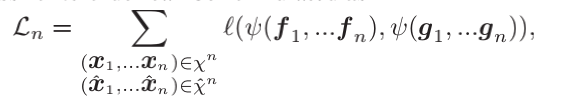
**Method**



**Framework**

通常，我们关注的信息阶数越高，可以获得的信息越多。受此启发，我们专注于信息关注的不同阶数，并提出了通过两阶段的知识迁移来促进低分辨率图像识别的混合阶数关系知识蒸馏（HORKD）（见图2）。 第一个跨结构蒸馏阶段将各种阶数关系知识从高分辨率的繁琐模型（**教师**）转移到模仿模型的行为的高分辨率紧凑模型（**助教**）。然后，我们将各种阶数关系知识从助教模型转移到一个低分辨率的紧凑模型（**学生**），该模型模仿跨分辨率蒸馏助手的行为。

**传统知识蒸馏方法**更加关注表示**特征空间中实例之间的逐点关系**，在该空间中**知识的传递可能不充分**。相比之下，我们的方法旨在通过**考虑高阶关系来实现更好的知识传递**。通过重新定义损失函数，学生可以很好地学习老师提取的结构知识，并有效地弥补了分辨率降低带来的必要信息的不足 ，从而提高其识别性能。为此，分别将χn和χˆn表示为一组不同的高分辨率和低分辨率实例的n阶元组，fi =φt（xi; wt）作为从高分辨率图像xi提取的教师知识 并且gi =φs（xˆi; ws）作为来自响应低分辨率图像xˆi的学生知识，可以将n阶蒸馏过程表示为：



其中ψ是一个**关系势函数**，用于测量给定n元组的关系能量，并且是一种损失，惩罚了师生之间的差异。

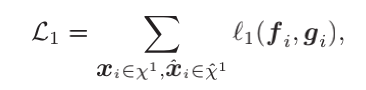
**Hybrid Order Relational Knowledge**

显然，**关系势函数ψ**在关系知识蒸馏中起着关键作用，它**影响知识蒸馏过程的有效性和效率**。通常，在花费更多计算时，**一个高阶势函数在捕获高层结构信息方面可能更强大**。 假设一个输出表示空间（例如mini batch）有m个示例，则n元组空间的大小为其组合数Cnm。

我们的蒸馏过程试图使教师流和学生流之间的**势能信息匹配**。 因此，**small batch normalization非常有用**，尤其是当两个流之间的差异很大时。为了补偿分辨率降低中的信息丢失，我们**希望传递各种阶数关系知识**。 因此，需要开发一种有效的解决方案以有效地解决这一问题。 在这项工作中，我们**首先采用传统的低阶关系知识：1阶，2阶和3阶**。 此外，我们提出了一种有效的方法来利用高阶相关知识。

**1-order relational knowledge.**

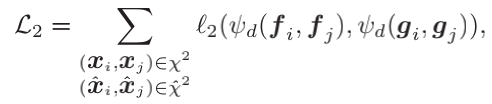
当n = 1且该关系为一元时，也称为**逐点蒸馏损失**。它在以前的作品中很流行，该作品将**老师产生的类概率**用作训练学生或**传递中间特征图**的软目标。 在这项工作中，我们转移特征，损失函数如下所示：



其中l1是一个测量**师生特征实例之间的L1距离**的函数。

**2-order relational knowledge.**

也称为**逐对蒸馏损失**。最近的工作已将其用于各种任务，例如图像分类，图像检索和语义分割。其目的是在实例之间**传递成对关系**，特别是在我们的方法中**成对相似性**。我们采用平方差来表示二阶关系知识蒸馏损失：

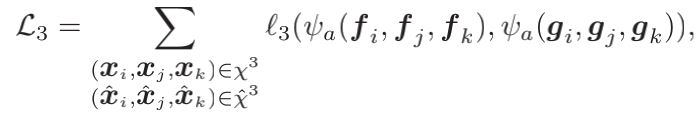


其中**ψd**是归一化的损失函数，用于测量mini-batch空间中**两个实例的特征之间的L2距离**

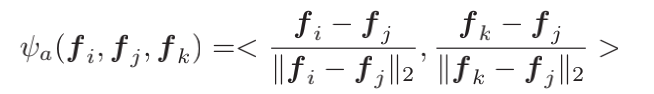
有，

**3-order relational knowledge**

它以三元组形式度量示例之间的关系。 为此，Park等。 提出一个角度蒸馏损失，该损失由输出特征空间中的三个示例形成：



其中l3是**Huber损失**。 角度（angle-wise）势函数**ψa**表示为：



其中<... ，... >是点积运算符。三阶关系知识通过**惩罚角度差**来传递嵌入训练实例的关系，由于三阶关系知识的阶数较高，因此可能能够传递更多有效信息，从而在训练学生方面更具灵活性。

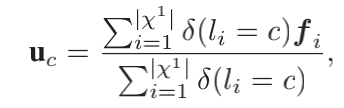
**Center-based relational knowledge**

通常，当**阶数大于3时**，它将带来以下两个问题：

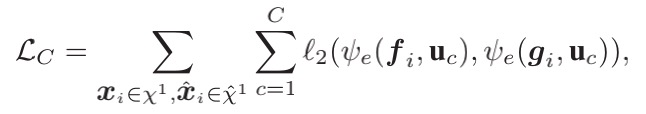
1. 正如我们期望的那样，**计算成本将增加**，
2. 势函数难以定义。

解决第一个问题的可行方法是减少训练时每个mini batch中的示例数。但是，很明显，示例与mini batch之外的其他一些示例之间的关系将丢失，从而**降低了知识转移的充分性**。 在这项工作中，我们**提出了以类为中心的关系知识来解决这些问题**。

一个额外的例子集，用来描述输出特征空间的类中心：



其中δ（li = c）是一个指标函数，如果li = c等于1，否则等于0。 C是总类数。这样，每个示例的特征在于特定类中心的特征实例，而类中心则由平均特征实例表示。 然后，**使用这些类中心构建一个支持空间**，该支持空间可用于创建实例的C阶关系知识：



其中**ψe**度量两个实例特征之间的**L2距离**，其中。 以此方式，该高阶关系可以被转换为一组2阶关系，其可以被有效地解决。

**Total distillation loss.**

通过混合阶数关系，要传递的知识可以涉及各种级别的关系，即个人级别，配对级别，三重态级别和group级别的知识。最后，总蒸馏损失是这四个损失的加权总和：



**Two-Stage Knowledge Transfer**

如图2所示，知识转移包括两个阶段。在**第一个跨结构蒸馏阶段**，教师和助手网络的输入是高分辨率图像。目的是**减轻结构冗余**（例如许多知识提炼方法的一般过程（Hinton，Vinyals和Dean 2015； Romero等人，2015））。

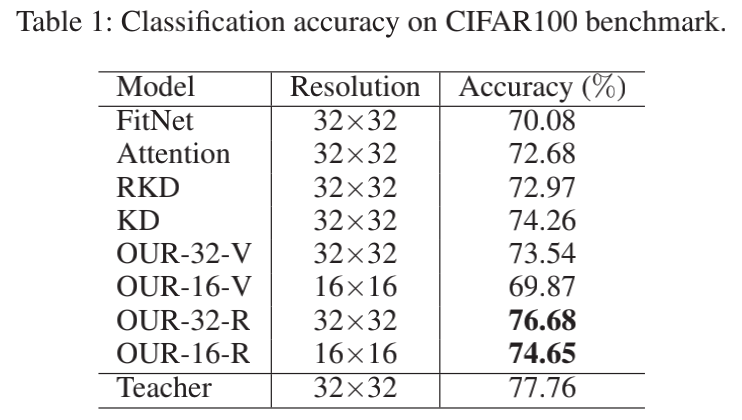
经过跨结构蒸馏后，**分辨率仍然存在冗余**，这意味着在充分了解相应的高分辨率图像后，仍可以识别出低分辨率图像。在第二个跨分辨率的步骤中，我们使用助理网络作为新老师来指导与助理网络具有相同结构的低分辨率学生网络的训练。目的是减少由于分辨率降低而导致的信息丢失。

**Experiments**

**Low-resolution Image Classifification**

数据集：CIFAR100

模型：ResNet50（32×32）ResNet18（16×16）



32-分辨率-Resnet/VGGNet

**Low-resolution Face Recognition**

老师：VGGFace2 学生：ResNet34

数据集：UMDFaces 验证集：LFW UCCS

