指导教师：杨涛　　提交时间：2016/3/11

CVPR2015 Paper Translation

No: 01

姓名：杨添文

学号：2013302534

班号：10011303

扩展目标检测器的视野：
视频中目标检测的增量学习框架

摘要：

在过去的几年中，它已被证明，基于图像的对象检测器对训练数据来说是敏感的，并且经常无法概括事例而超出原训练样本域（例如视频）。人们已经提出一些域自适应技术来解决这个问题，利用 DA 技术设计一个固定的复杂模型适应于新的领域（例如，视频）。我们假定无标号数据不应只允许适应,而且能提高(或至少保持)性能原始和其他领域通过动态调整模型的复杂性和参数。我们称之为概念域的扩张。为此,我们开发一个新的可伸缩的和准确的增量的目标检测算法,基于几种扩展的优势嵌入(LME)。我们的检测模型包含一个嵌入空间和多个类原型的嵌入空间,代表对象类;这些原型的差异让我们思考利用多检测。通过逐步检测对象实例的视频并增添了对模型检测的信心,我们能够动态调整的复杂性的探测器随着时间的推移,实例化新的原型跨越到所有域模型。我们测试性能的方法是通过扩展一个在 ImageNet（ImageNet 是一个计算机视觉系统识别项目， 是目前世界上图像识别最大的数据库）上实例化后的对象探测器，来检测以自我为中心的视频的日常生活活动(ADL)数据集的对象和挑战来自 YouTube 视频对象的数据集。

1、介绍：

在过去的几年中，已证明在对象检测数据集中存在着重要的差异，以及在这样的数据集和真实世界的图像之间。结果,监督分类器/检测器对准一个数据集,往往不能充分地在另一个,或现实世界图像中工作,统计数据可能没有被捕获在最初(标记)的训练数据集。处理这种偏见的监督,半监督和管理领域提出了适应方法,分类和实值回归任务。

虽然大多数领域适应技术专注于应用程序的培训和测试实例图片,利用传统相机,解决图像到视频对象探测自适应的问题。在 Image-to-video 场景是既引人注目又有挑战性。利用图像数据集训练非常值得在视频探测器中使用,因为图像容易标签和大量丰富的标签数据集已经存在。获得视频相当于 ImageNet，对范围而言,将是一个不可

逾越的任务。然而,通常存在显著的外观图像之间的差异(例如,在 web 上获得)和视频(例如,在 Youtube 上获得或使用自己的相机)。网页的图片往往是高分辨率的对象中心。视频,另一方面,通常处在低分辨率,不以对象为中心,至少在自我中心的设置有广泛不同的外观，因为传感器和运动工件的质量。因此，域图像和视频之间的转变往往是严重的。

几乎所有领域适应技术假定数据分为离散定义域,通常一个源(训练)域目标(测试)域,和任务是有效地传输信息(或标签样本)从源到目标域。离散域的概念,只关注性能的目标域有点太过于简单化了。实际上,如上所述,目标域通常是不断演进的。此外,你可以认为,作为对象实例,外表,照明和观点正在发生变化,由此产生的进化是一个扩张的原始域对象,形成一个新的或旧的进化领域。这两者之间的差别非常微小。在连续域适应(和增量学习)我们的目标是不断适应(或学习)固定的复杂模型，进行尽可能准确地到达目标批数据。我们主张不断调整模型本身的复杂性。这应该允许调整模型不仅提高对到达的数据,但也至少保持其性能之前和未来的领域。我们也不假定数据达到一个不断变化的流。

为此我们提出一个增量,自我启发,方法扩展域的图像到未标记视频。我们从一开始大幅度嵌入(LME)模型,这是我们适应的检测任务。使用这种检测模型,找到标记到视频中的对象,提取和跟踪最有把握的实例。如果从这些痕迹形成一个集群实例,它们进一步用于调整模型的复杂性，通过添加新类原型。这个过程提取的实例和学习扩展域模型,继续作为额外的视频。

我们的方法是受终身学习的首要目标的鼓舞。我们注意到,我们的方法是相关的次范畴化,但又不像次范畴化,这是假设完全标记或略微标记的实例,我们在一个完全无监督的情况下工作。此外,虽然子类别,一般来说,在外表上的空间不形成任何形式的相干结构,我们的模型可以确保对象类原型形成一个连贯的歧管,通过正则化,在学习上限制偏移。

贡献：我们的主要贡献是增量的框架领域扩张,在一个基于图像的目标检测模型的复杂性不断适应新到达的无标号视频的方式,随着时间的推移,改善了性能在视频领域的发展，但同时保持原始图像域上(或提高)的准确性。有效的领域扩张,是关乎到建立一个更好的探测器使用无监督视频数据的。作为这个大目标的一部分,我们制

定一个新的目标检测模型,灵感来自于优势嵌入(LME)。我们展示如何扩展LME 从多层次对象分类多目标检测问题,通过引入新的检测约束处理消极的实例。我们也为 LME 提出一个概率公式,允许模型执行直观信心评估测试实例,支持增量的学习。我们展示了增量域扩张有效应用对象探测器,它的训练只有 ImageNet 到视频,可以提高性能48%(13%通过扩张)对原始 LME ADL 数据集和 15%人数据集。

2、相关著作：

从图像到图像域的域适应：我们的领域扩张方法与领域适应(DA)密切相关,这是一个统计方法,侧重于适应现有的模型在一个域(源)到一个新的数据域(目标)。领域适应气候变化可以分为监督方法和无监督方法,标签可用于样品在目标域,没有提供标签的地方。我们的方法在后者的情况下,作为我们的目标则是扩大模型学习标记图像，其中包含标记视频数据。方法余现有方法之间的主要区别是,在现有方法假设存在多个离散域的情况下,我们认为相关的所有领域,如,模型的源和目标数据在一个连续的管汇两者之间没有明确的区别。基于这样的假设,我们的模型旨在改善在源和目标域,而大多数方法只关心给定目标域上的性能。

基于图像到视频训练的适应对象探测器：众多 DA 任务,以及适应探测器的任务训练从图像到未标记视频数据是一个特别感兴趣的话题,很大程度上是由于视频数据的注释。许多模型采取的战略选择的是正面和负面的样本测试数据，基于他们对现有的探测器的把握。在基线检测器与低阈值生成正/负样本的新词汇树分类器,然后决定在标签上。我们的模型还利用了现有的模型选择测试样本,但在我们的例子中,模型是不固定的。我们也可以扩大不仅旨在改善在视频领域,而且维护或提高,性能检测的图像域。在决定加入到训练池,许多作品进一步利用视频帧的时间连续性数据票数等利用跟踪、匹配追踪，从一个基线检测探测器作为额外的阳性样本。

唐家璇等提出了一个自学的方法,逐步增加积极的样本,提高分类性能。与我们自学学习算法是相似的,但我们扩大模型而不是培训。多纳休等提出了一种方法,结合一个实例相似图形调整模型,并应用到视频数据的情况下,实例在一个轨道的距离被利用作为辅助实例的相似性。我们的方法还利用这些实体之间的相似之处,但它的模型

组(视频)种类相似,相似图不是给定的,而是含蓄地由视频的顺序到达。有些作品企图反面使用弱监督 YouTube 视频训练图像对象探测器。

自主学习：我们排列无标号的视频样本，根据分类,自主学习有关,并给出了数据点一个有意义的秩序,这通常是给给定样本分类的困难。自主学习被用来学习潜变量,它被用来发现对象类别从集群图像补丁。自主学习也用于唐，逐步添加标记样本的标记池。我们的工作利用类似的选择方法,但我们的模型考虑多级认为股份时模型。

终身学习：终身学习的理念,这是一个多屏画面等等学习转移在早些时候学到的知识学习阶段后期,首次构想,并在永无止境的成功语言学习者(内尔)已成为一个活跃的研究主题。内尔是一种增量式模型,了解新概念和规则，通过不断观察文本输入。类似的工作已经提出了图像数据的情况。我们的工作也可以看作是终身学习的一个实例,由于模型不断改进利用源源不断的输入,和新子类的原型的上下文中学习现有的类别原型。

大幅度的流形嵌入模型识别：我们的模型建立在优势(类)嵌入(LME),旨在学习空间优化的阶级辨别。LME 是最近流行的,很大程度上是由于它的能力扩展到许多类标签,与图像分类变得越来越重要越来越专注于大规模数据集。虽然有很多变异 LME,特别相关的是它提出了一个概率 multi-centroid 模型,就是与我们的相似。然而,每个类的 k-centroids 从 k - means 聚类得到原来的标签样本,而在我们的模型中多个重心逐步添加模型扩展的新视频。

3、增量学习框架：

我们考虑的一般问题应用对象探测器,训练图像,检测视频中的对象是完全无监督的方式。

一旦最初的 LME 检测模型经过训练，我们就可以利用未标记的数据序列的视频不断提升学习模型。提出一种增量式迭代学习框架,改进和增加了模型的复杂性是必要的,包括以下步骤:

1、从每个视频我们提取对象建议{bi}N="1,使用[35],和相应的特征向量{xjN^i}。

2、我们评估每个 Xi 使用提出的概率多中心 LME 模型来获得一组检测、标签 V ={xi, ci,p(y = ci,d = 1|xi)}Nl"i 。

3、我们延长组检测，利用时态一致性。

4、最后,我们使用选定的样本,更

新模型描述。

3.1、背景：大边缘嵌入

大幅度的嵌入方法分类,项目样品到低维空间,达到分离属于不同的类的实例中,对欧几里得度量。

让我们表示 d(子、加州大学)投影特征向量之间的相似度测量子= Wxi 和一个原型。LME 上面描述的目标可以通过积极的保证金编码子之间的相似性及其真正的原型和其他所有的原型:

$$d(Zi, Uy.) + \ ^{ic} > d(Zi, Uc) + \ 1,$$

$$i = \{1 \ \cdots N\}, c = \{1 \ldots C\}, c = yi,$$

扮演角色的松弛变量,我们想最小化。最优的学习{ ui,…,可以制定为最小化:}

4、对象探测的多原型：

最初的LME模型是专为对象分类的任务。我们扩展 LME 配方，使其适用于目标检测,并提供相应的概率解释。我们也推出多原型配方和现在的增量学习算法。

4.1、对象探测的 LME 模型：

微不足道的方式，扩展了 LME 对象检测模型来假设存在一个非类。然而,这将导致 LME 建模的非类使用一个非原型。因为外表的变化在非类是更高的然后在任何其他类,这可能不是最理想的。

我们注意到特定的相似性测量，这实际上推动负样本的中心嵌入空间和有效的特征选择(或抑制)之间所有的积极和消极类;相对于其他指标,例如,一个欧式度量,几何解释是不同的。

数值优化：EQ 的优化，使用交替优化,我们交替求解 U 和 W,同时保持其他变量固定不变,使用随机梯度下降。重复交替过程直到收敛性判据是 M2。

4.2、概率 LME 解释：

对估计探测器的把握将在排序和选择样本的关键域扩张。我们损失一部分,然而,处理的是一个多层次的问题,所以我们得到以下 LME 的概率解释。我们定义一个样本的后验概率,被认为是检测属于丙类,通过映射之间的相似性预测实例和类嵌入在 0 和 1 之间范围。

4.3、LME 多原型：

域的转变是伴随着在原始空间变化特性分布,因此,在 lowdimensional 中嵌入空间。这种转变导致探测器的性能降低和应对等领域的转变,我们需要一个更灵活的类表示在嵌入空间。我们学习一些,Kc,原型为每个类 c 代表多通道特性分布跨领域。然后,一个类和

实例之间的相似性得分可以计算使用同一个类的不同原型的相似之处。

参数越大,越接近 max()的函数。多原型模型的优化问题可以以同样的方式制定 LME 模型检测约束 Eq。

4.4、增量多原型 LME 模型扩张：

多原型 LME 模型是自然适合域的扩张。遇到新的数据模型,由目前的原型,不能很好地近似，我们可以逐步添加新的原型更精确地模型嵌入空间的分布特征。学习一门新原型的问题才能在 LME 框架内制定增量学习过程。

假设我们想扩大基于原型的表示为类 cn。当添加一个新的原型模型应该满足两个属性:(1)的新原型应该代表和偏移类;(2)不应该引起其他类样本的误分类,也就是说,它应该足够远离现有类别原型为其他类。更正式,可以制定如下优化问题:

W 是新学到的数据嵌入,我们现有的数据嵌入,情况是对于给定的类别,最初的原型和 Ucn =[Ucn,Ucn]。(14)是一个LME约束之间的新类别和现有类别。(15)之间是相同的约束每一个新类别嵌入现有的类别。(16)是检测约束。参数 v 和 n 是正规化确定类似新学会了嵌入的应该是最初的类别和数据嵌入。优化问题在 Eq(13)-(16)非凸,但提供了一个好的初始化随机梯度下降法可以获得合理的局部最小值。

4.5、从未标记的视频中发现对象：

给定一个未标记的视频,我们首先提取检测，通过计算对象的初始方案及其使用特性形成现成的建议方法。然后我们提取视觉特征，为每个对象提议我和评估他们使用 Xi 多原型模型获得概率得分。然后一组检测对象 Vv = { Xi,ci,p(y = ci,d = 1 \ Xi)}可以建议我选择对象,形成的表面张力 p(y = ci,d = 1 \ Xi)> v,v 是一些阈值;v = 0.6 允许我们在我们的实验中获得相当不错的结果。获得的一组检测，可以使用新的积极训练样本训练新类别原型。

轨道的形成：获得更多的样品：我们进一步利用时态一致性,即,如果检测到对象在一个框架,它可能持续的帧数在相对相似的位置和规模。

具体来说,我们使用想法[31]在试验的基础上,提出了从视频中提取跟踪使用 KLT 追踪。后一组计算对象建议 D 与相应的边界框{ bi } i = 1 l,每个对象建议边界框 bi,我们选择相交的最长追踪。然后计算物体的相对位置，建议相交这个跨帧跟踪,在每一帧选择最高的提议帕斯卡和 bi 席卷轨道重叠。用这种方法得到一组对象建议每个 bi

构成轨道。获得跟踪分数,我们评价他们,接受跟踪如果超过一半的检测在跑道上，有 p(y ^ = c,D = 1 \ x)如果跟踪被接受,我们将所有的样品跟踪到 D。

5、尝试：

我们在现实世界的图像和视频数据集上验证我们的方法,图像数据集,训练基地探测器,我们使用的子集 ImageNet 数据集与相应的课程。我们使用的不相交的子集 ImageNet 检测器的性能，报告图片之前和之后在适当的地方的增量域扩张。我们测试方法的日常生活活动(ADL)和 YouTube 对象(人)数据集:

分布数据集：ADL 数据集包含 20 个第一人的视频,记录下不同的主题。这是一个具有挑战性的数据集和探测器的直接应用于静态图像,因为对象遭受大的观点/规模变化和遮挡,因为每个视频交互。ADL 数据集的边界框注释对象有 48 类。我们选择 8 类最常见的一个子集,即瓶,冰箱,微波炉,杯子,烤箱,肥皂液,水龙头,和电视,来测试我们的模型。

人数据集：网络视频,每个视频包含一个对象的类:飞机,鸟,船,汽车,猫,牛,狗,马,摩托车,火车。数据集分为训练集和测试部分,在试验部分包含一个单帧边界框注释的目标对象评估/视频。我们使用测试数据集的一部分;它包含 15 - 60 为每个类视频。

5.1、定量的评价：

我们评估的目标检测性能基线和模型使用意味着平均(mAP)[7]onADL(表 1)和人(表二)数据集。

	bottle	fridge	microwave	mug/cup	oven/stove	soap liquid	tap	tv	av. ADL	ImageNet
DPM[25]	**9.8**	0.4	20.2	14.8	0.1	**2.5**	0.1	26.9	9.35	–
GK [12]	2.11	1.77	41.19	14.70	19.57	0.20	1.62	**60.67**	17.73	–
LME	0.00	0.28	3.07	0.00	0.52	0.03	0.55	3.73	1.02	–
LME-A	1.93	**3.42**	40.30	18.34	27.84	0.37	1.46	53.26	18.36	76.96
LME-D	1.69	1.63	39.87	13.06	19.33	0.35	1.67	40.64	14.78	78.91
LME-DT	1.85	1.76	52.37	15.91	24.54	0.42	**2.41**	56.16	19.43	–
IDE-LME	2.04	2.73	**56.69**	**21.86**	**29.94**	0.25	2.26	59.53	**21.91**	**79.23**

表 1。检测性能平均每个类和所有类别的地图 ADL 数据集,基线和我们的方法的变体。我们报告检测结果 ImageNet 子集包含 8 类 ADL 数据集。

图 2。说明数据和检测结果。上面一行:ImageNet 图像的例子,用于训练的初始模型。底下一行:实例使用 IDE-LME 正确检测的检测。注意的显著差异出现在 ImageNet 和 ADL 数据集对象。

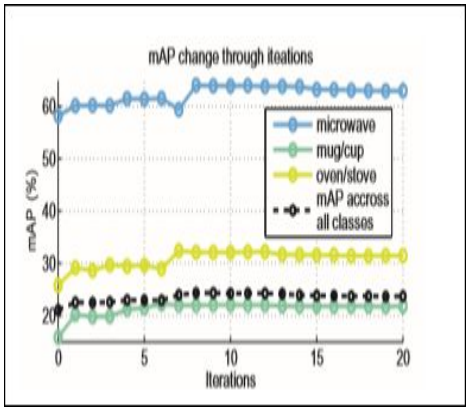

图 3。地图作为视频看到的函数(x 轴)的子集类分布数据集;用于扩张的地图是整个视频平均反诽谤联盟和其他视频数据集;在看不见的视频中性能提高了。

训练图像因为没有检测约束，所以执行很差,增加了检测性能约束的结果与 DPM 基线持平。将时间一致性利用跟踪(LME-DT)，性能提高了 31%以上，以及对 LME-D ADL 数据集人的数据集(5%)。增量更新模型(IDE-LME)使用我们的方法进一步带来了显著的性能提高，YTO 数据集可以归因于这样一个事实:人数据集，包含一个或几个对象在典型的观点,另一个原因是稀疏的人数据集的注释,限制的能力估计的性能提升。

注意到其他基线,IDE-LME 在类初始精度高(如电视和微波 ADL 数据集),而对其他类执行明显恶化。我们相信,这些课程有效地确定 GK 变换,而其他类的分布的变化是不考虑的。另一个趋势出现在两个数据集，初始模型应该有足够的精度能够让选择的视频样本有效的工作,否则会发生轻微的性能下降(液体类人数据集或猫类的数据集)。

以上实验表明模型的复杂性增加了前所未有的对象类的外观的变化。支持这种说法也表明,我们的模型改善了原始图像域,我们报告的分类结果测试 ImageNet 数据集的分割。这表明,新添加的样品不仅提高检测性能的测试视频数据,而且还在源图像数据上提高分类的性能。注意定义域适应(DA)基线 LME-A，他改善视频但是在两个数据集降解上对源图像域(DA)是一个典型的行为。

模型的性能通常增加更多的观察视频,但经过渐近线前 10 的迭代后，ADL 数据集(参见图 3)。这种早期性能饱和可能是因为高外观对象之间的相似性在目标域(自我中心的视频)。人数据集显示了类似的趋势。然而,如果目标领域不断变化或发展，随着时间的推移,性能可能会继续增加。

	aero	bird	boat	car	cat	cow	dog	horse	mbike	train	av. YTO	ImageNet
DPM[18]	30.79	10.46	0.97	48.62	18.30	33.69	13.67	**26.78**	35.85	**23.98**	24.31	–
GK[12]	**40.05**	23.16	24.44	32.62	**24.26**	38.26	24.23	17.75	**36.27**	10.69	27.17	–
LME	35.00	24.13	16.08	27.41	4.30	31.18	2.12	0.23	6.83	10.30	15.75	–
LME-A	39.78	**35.18**	35.20	**48.67**	15.02	37.90	**30.70**	25.86	28.93	10.82	**30.80**	79.91
LME-D	29.61	22.91	32.39	25.53	18.63	38.94	15.55	9.22	31.47	12.09	23.63	83.16
LME-DT	31.67	21.83	40.13	25.94	17.59	41.44	15.47	11.74	30.56	13.67	25.00	–
IDE-LME	33.07	21.40	**42.26**	34.49	18.33	**46.92**	17.24	11.83	34.73	12.50	**27.28**	**83.20**

表 2。检测性能为每个类和平均映射(mAP)在所有类在 YouTube 对象(人)的数据集。

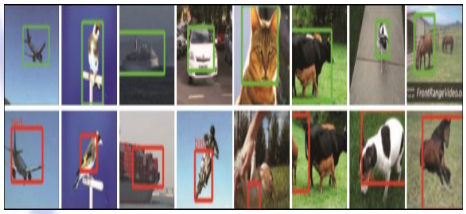

图 4。上面一行:正确地检测到对象的实例。底下一行:误检测的例子，由于不正确的分类,不准确的边界框,或不正确的标签。

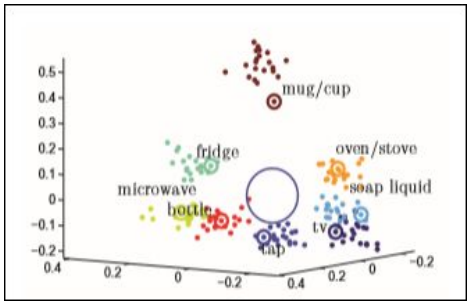

图 5。可视化学习(扩大)多中心 LME,ADL 数据集,投射到 3D 空间;初始原型都有额外的圆。

5.2、定量的分析：

图 2 展示的例子检测 ADL 数据集。注意到源域和目标域之间的显著差异。图 5 的三维可视化是学会了八维嵌入,在每个类别都被表示为一组类别原型扩大学习的过程。我们观察到某些对象类,比如杯子,后来添加的原型放置远离原来的中心,代表特性分布变化之间的杯子类 ImageNet 数据集和 ADL 数据集。

图 4 显示了检测人数据集的例子,使用 IDE-LME 获得。注意,对象通常是确定正确,但边界框太小或太大。我们把这种现象归因于背景包括大部分 ImageNet 图像,这可能是因为松检测高于紧的。

6、结论：

在本文中,我们解决了扩张域的问题,在对象的范围探测器中得知在初始图像标记逐步扩展到覆盖未标记的视频。为此,我们已经开发出一种以新颖的网络概率多中心的大边缘嵌入与检测约束模型,其中每个对象类别与多个原型,代表数量逐步增加,学习算法选择有把握传入的未标记的样本数据。试验中 ADL 和人的公共数据表明,该模型不仅大大提高了检测目标标记的视频,还有源图像域的性能。增量域扩张模型可以作为终身学习对象侦测系统模型，并且可以扩展到包含标记新视频数据的连续流。一个可能出现的潜在问题是漂移模型，我们没有见过这个实验和正则化是为了防止这一点,但它是可能的,这种漂移可能出现规模大得多的数据集。对于未来的工作,我们计划去探索人类循环系统与主动学习来防止这样的漂移,使得模型能够基本上自身学习,并且很少人工干预仅由模型的要求自定计划。

7、参考文献：

[1] Y. Aytar and A. Zisserman. Tabula rasa: Model transfer for object category detection. In ICCV, 2011.

[2] S. Bengio, J. Weston, and D. Grangier. Label embedding trees for large multi-class task. In NIPS, 2010.

[3] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka,andT.M.Mitchell.Towardanarchitectureforn everending language learning. In AAAI, 2010.

[4] X. Chen, A. Shrivastava, and A. Gupta. Neil: Extracting visual knowledge from web data. In ICCV, 2013.

[5] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. F.-F. ei. Imagenet: A Large-Scale Hierarchical Image Database. In CVPR, 2009.

[6] J. Donahue, J. Hoffman, E. Rodner, K. Saenko, and T. Darrell. Semi-supervised domain adaptation with instance constraints. In CVPR, 2013.

[7] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. IJCV, June 2010.

[8] P.F.Felzenszwalb,R.B.Girshick,D.McAllester,andD.Ra manan. Object detection with discriminatively trained partbased models. TPAMI, Sept. 2010.

[9] B.Fernando,A.Habrard,M.Sebban,andT.Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In ICCV, 2013.

[10] Y. Fu, T. Hospedales, T. Xiang, Z. Fu, and S. Gong. Transductive multi-view embedding for zero-shot recognition and annotation. In ECCV, 2014.

[11] A. Gaidon, G. Zen, and J. A. Rodriguez-Serrano. Selflearningcamera: Autonomousadaptationofobjectdetectors to unlabeled video streams. In Arxiv, 2014.

[12] B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic fl ow kernel for unsupervised domain adaptation. In CVPR, 2012.

[13] R. Gopalan, R. Li, and R. Chellappa. Domain adaptation for object recognition: An unsupervised approach. In ICCV, 2011.

[14] M. Hoai and A. Zisserman. Discriminative subcategorization. In CVPR, 2013.

[15] J. Hoffman, T. Darrell, and K. Saenko. Continuous manifoldbasedadaptationforevolvingvisualdomains. InCVPR, 2014.

[16]J.Hoffman,B.Kulis,T.Darrell,andK.Saenko. Discovering latentdomainsformultisourcedomainadaptation. InECCV, 2012.

[17] Y.Jia,E.Shelhamer,J.Donahue,S.Karayev,J.Long,R.Gir shick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. Arxiv, 2014.

[18] V.Kalogeiton,V.Ferrari,andC.Schmid. Analysingdomain shiftfactorsbetweenvideosandimagesforobjectdetectio n. Arxiv, 2015.

[19] A.Khosla,T.Zhou,T.Malisiewicz,A.A.Efros,andA.Torr alba. Undoing the damage of dataset bias. In ECCV, 2012.

[20] B. Kulis, K. Saenko, and T. Darrell. What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In CVPR, 2011.

[21] M. P. Kumar, B. Packer, and D. Koller. Self-paced learning for latent variable models. In NIPS, 2010.

[22] Y. J. Lee and K. Grauman. Learning the easy things fi rst: Self-paced visual category discovery. In CVPR, 2011.

[23] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In International Joint Conference on Arti ficial Intelligence (IJCAI), 1981.

[24] T. Mensink, J. Verbeek, F. Perronnin, and G. Csurka. Distance-based image classi fi cation: Generalizing to new classes at near-zero cost. TPAMI, 2013.

[25] H. Pirsiavash and D. Ramanan. Detecting activities of daily living in fi rst-person camera views. In CVPR, 2012.

[26] A. Prest, C. Leistner, J. Civera, C. Schmid, and V. Ferrari. Learning object class detectors from weakly annotated video. In CVPR, June 2012.

[27] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In ECCV, 2010.

[28] P. Sharma, C. Huang, and R. Nevatia. Unsupervised incremental learning for improved object detection in a video. In CVPR, 2012.

[29] P. Sharma and R. Nevatia. Efficient detector adaptation for object detection in video. In CVPR, 2013.

[30] D. L. Silver, Q. Yang, and L. Li. Lifelong machine learning systems: Beyond learning algorithms. In AAAI, 2013.

[31] K. Tang, V. Ramanathan, L. Fei-Fei, and D. Koller. Shifting weights: Adapting object detectors from image to video. In NIPS, 2012.

[32] S. Thrun. A lifelong learning perspective for mobile robot control. In Intelligent Robots and Systems. 1995.

[33] C. Tomasi and T. Kanade. Detection and tracking of point features. Technical report, 1991.

[34] A. Torralba and A. A. Efros. Unbiased look at dataset bias. In CVPR, 2011.

[35] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. IJCV, 2013.

[36] X.Wang,G.Hua,andT.Han. Detectionbydetections: Nonparametric detector adaptation for a video. In CVPR, June 2012.

[37] K. Q. Weinberger and O. Chapelle. Large margin taxonomy embedding for document categorization. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, NIPS. 2009.

[38] J. Weston, S. Bengio, and N. Usunier. Wsabie: Scaling up to large vocabulary image annotation. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 2011.

[39] M. Yamada, Y. Chang, and L. Sigal. Domain adaptation for structured regression,. In IJCV, 2014.