

指导教师： 杨涛

提交时间： 2013/3/19

CVPR2015 Paper Translation

No: 01

姓名： 王文新

学号： 2013302531

班号： 10011303

成为专家——互动多机教学

爱德华 约翰 Oisín Mac Aodha 加布里埃尔*j *brostow

摘要

与机器相比，人类是极其擅长分类图像，尤其是他们具有先验知识的类别。如果这之前的信息不可用，监督形式的教学图片是必需的。为了更快的学习分类，人们应该看到一些重要的和典型的图像，其次是不太重要的图像或者不是所有的图像。然而，图像重要性是细化的，一个教学图像对一个学生是重要的如果它改变了学生们的区分不同课程的总体的能力。更远了说，学生一直学习，然而图像重要性依赖与他们现有的知识，而且随之时间变化。

在这份工作中，我们建议一种交互式机器学习算法可以将具有挑战性的视觉概念教给人们。我们的交互式算法选择在线的具有图像标签的教学设置可以向学生们展示它们所学的东西。我们展示了一种策略，就是一种概率可以模型出学生的能力和进步，依据他们正确的和错误的回答。我们用真实的参与者穿过多样的和挑战性的数据集。

1 介绍

大的，手动的，带注释的图像数据集对于最近的增长做出了贡献与核心的电脑视觉问题例如对象检测和分类【16. 37. 25】。假若兴趣的视觉分类是在每天所见的物体，带注释的可以被完成被拥挤的源头标签从互联网使用服务例如亚马逊土耳其机器人。一个典型的图像被一个结构集贴标签，从阶级利益的例子图像。这个注解者分配累标签的新图像在地面真理是未知的。

但如果他们不确定会发生什么？这是当代码是错误的现实问题的假设

对有兴趣的课程有优先认识人类在计算机教学中，计算机，而不是人，是（完美）的甲骨文，并负责提供一个教给学生的教学，帮助他们学习更有效的任务。教人一个新的技能是有用的，在它自己的权利。此外，他们现在更好定位准确注释附加原始教学集外的数据。自动教学算法在许多领域都有广泛的应用，语言学习，医学图像分析，生物物种鉴定[30]，以及更多。至关重要的是，自动教学是有效的，它需要能够评估学生目前的知识，并有一个机制为选择教学实例，最好提高这个知识。在这项工作中，我们专注于图像分类的任务。在这里，这是不可能的教师直接‘教’的高维决策边界的人学习者，因此，学生必须学习这个边界通过展示教学图像。我们的教学目标是选择教学图像，将最大限度地提高学生的分类能力的最小量教学时间。与计算机不同，人类既有有限的和不完美的记忆，例如水平识别，特别是在最初的学习任务[17]。不过，人类拥有的优势是拥有能力推广到未知的例子，并进行域自适应给出的实例。以前的大多数在机器教学中工作的重点是非互动性教学，一个教学集离线计算，独立每个学生的反馈[39, 33]。在这工作中，我们探讨了在探索问题的互动教学[14, 4]。在这里，老师可以适应他们的教学根据个人的当前性能设置在线学生（见图 1）。我们提出了一种算法，交互式地教多视觉范畴对人类学习者的学习。我们的贡献有三：1) 不同于电脑，人类是不是最优的学习者。我们的算法模型学生的能力在线，导致在教学上设置，适合每个人学生。我们不做任何关于内部的假设学生使用的学习模式。相反，我们目前他们

的教学图像，试图减少他们预测未来的不确定度的基础上的估计他们现在的知识。2) 我们的教学算法减少了学生学习分类的时间量涉及多个类的任务。实验证明真正的人类参与者，使用我们的算法，执行比一些具有挑战性的数据集的其他基线。3) 最后，我们提供了一个基于网络的界面和框架探索新的教学策略。我们的意图是这将鼓励发展新的和多样化的多种视觉学习的教学策略

任务。

一个更大的标记的图像图像数据集，同时隐藏真正的类标签。这个学习者/学生对图像的估计作出反应类。老师然后更新他们的学生模型，最后，揭示了正确的答案。这个过程在教学结束前，再重复使用图像。

他们可以概括，无论是从日常生活中或从专业培训。对于许多问题，高度专业化，特定领域的知识，通过广泛的获取训练，是需要别人能分辨之前潜在的多个高度自相似的对象类别教学图像或“教学集”的设计显示代码是具有挑战性的，因为每个注释将拥有不同程度的专业知识。虽然这是可能的模型的不确定性和噪声产生的组注解者提高他们的集体表演 [34, 40, 27]，这些方法往往 downweight 票从弱通过学习信任的专家注解。在本文中，我们提出问题-如何成为一个专家？我们假定一个给定的视觉人类的判别能力分类任务可以提高通过更好的建模教学过程中需要使他们的专家。

被称为机器教学的方法之家提供一般的解决问题的教学人类 [43, 18, 42, 39, 33]。机器教学不一样作为主动学习 [38]。在主动学习中，计算机的目标是要学习更多的精确模型，给出最小的监管金额。这是通过仔细选择只有最翔实的数据点被标记的人。

2. 相关工作

在这里我们最密切相关的工作在机器教学。正如我们所关心的任务的图像分类，我们专注于教学分类的研究功能。然而，值得注意的是，不同的文献中已经探讨了教学任务的类型，例如，连续的决策任务 [8]。人类是如何获得的代表类别是视觉研究的一个活跃领域心理学。范畴获取的许多候选模型在人类的存在，并为一个概述我们的读者直接到 [28, 35]。在这项工作中，我们的目标是不要直接对这些内部流程进行建模，而是要对它们进行建模把人当作一个随机的黑盒子学习者。为了方便，我们将相关工作分为机械教学双区-批（固定）教学和交互式（自适应或在线）教学。对于最近的，和一般的介绍对机器教学，请参阅 [43]。

机器教学-批（固定）在批量教学中，教师的目标是构建离线的一组最佳的教学实例，这是

在教学中呈现给学生。早期的工作这一领域的重点放在教学的理论分析上尺寸 [18]。教学维度被定义为给定的实例所需的最小数目概念教学的概念，一个学生。像许多其他在教学中， [18]做了简化的假设那个学生有一个完美的记忆（即一次展示一个例子学生会记住它的未来）-一个假设在现实世界教学中被侵犯。其他理论动机的作品，虽然有趣，提供小真实的人类受试者的验证 [3, 13, 46]。

最近，朱 [42] 试图最大限度地减少联合教师的努力与学生的优化直接在教学中设置。该

模型假设学生的学习模式是已知的老师，并且它是在指数家庭。在后续的工作中，Patil 等人。 [33] 维护与计算机不同，它具有无限的存储能力，人类的检索能力是有限的。动机通过真正的人类研究 [17]，他们表明，建模有限的的能力，提高了人类的学习成绩简单的一维刺激的任务。最相关的工作，该等。 [39] 教二元视觉概

念，展示图像，真正的人类学习者。他们的方法操作离线，并试图找到一套教学实例，最好传达一个已知的线性分类边界。机械且实验显示出改善相比其他基线，包括随机抽样。他们的方法尝试编码一些噪声容入教学集，但仍无法在教学过程中适应学生的反应，因为教学图像的顺序是固定的。

人机交互（自适应）

真正的人的学生往往是噪音，特别是在学习的早期阶段，当学习的概念不是在他们的头脑中形成的。此外，学生不都以同样的速度学习——对一些有困难的概念学生可能会更容易为他人。在互动教学中（图1），老师收到来自学生的反馈信息作为教学进展。鉴于此反馈，教学策略能适应一个个体学生的当前能力随着时间的推移。利用学生的概率模型和无噪声学习的假设，都和凌[14]提出了一个教学策略称为“最坏的预测”。这种策略是相似的

对不确定抽样，一般在主动学习[38]。然而，与积极的学习，在机器教学的老师有接地的真理类标签，可以使用此来评估学生的表现在教学过程中。实验，我们表明，他们策略执行次优，因为它只寻求显示学生的形象，他们目前最不确定关于，不考虑如何对图像的信息可能与他人的关系。因此，它很容易受到教学点，即不具有代表性的图像边缘教学设置。在一些互动式教学论文中与视觉概念，巴苏和克里斯坦森[4]评价二元分类中的人类学习性能三种不同的教学方法。学生的任务分类简单综合生成（和线性可分）描写蘑菇为两类。他们没有明确地从学生的标签噪声，但调查不同的界面设计和特色空间探索的方法，帮助教学的学生。在本文中，我们解决这个问题的互动多类真实图像的多类真实图像的建模学生在教学过程中提供反馈的能力。

3. 机器教学

在这一节中，我们正式定义我们的机器教学任务。我们的老师电脑有一个标记数据集 $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$ 每个希一种二维特征向量编码图像二，彝族 $\in \{1\} \dots, C$ 是其相应的类标签。这个教师的目标是“教”的分类任务的人学习者通过显示他们的图像从数据集。我们把这些教学图像作为教学组， DT ，的一个子集的图像从 D 那里 $|DT| \leq |D|$ 。在每一个一轮互动教学，教师首先选择一个形象由特征向量 x_t 表示展现给人类学习者。

老师向学生显示图像，它是不可能直接显示他们的高维特征向量 x_t 。图像显示的选择是基于一个过程，我们称之为“教学策略”，老师，用的。首先，老师只显示图像和不显示地面真相类标签。通过不透露这个类的标签，老师可以问学生要陈述自己的形象是属于哪一类的

去。收到学生的反应后，老师然后更新它的模型，然后揭示了地面真标号。教学所得为一组

教学轮，并在每次迭代中，教师获得更好地了解学生的当前能力。图1列出了一个教学迭代。接触地面的真理，老师很知道的条件分布 $P(y|x)$ 为每个数据点西安。学生学习者有相应的分配 $PL(y|x)$ ，仅基于训练的例子，他们所看到的到目前为止。在教学过程中，教师要尽量减少学生的期望损失 $EX = L(P(y|x), PL(y|x))$ ，

(1) 在数据集，其中 $l()$ 是一个适当的分类损失函数。然而，老师却没有办法直接观察学生真实的课堂条件分布， $PL(y|x)$ ，所以必须近似为 $\hat{PL}(y|x)$ 。在这纸，我们代表 $\hat{PL}(y|x)$ 使用概率，半监督，分类。

3.1. 教学策略

最佳的教学策略是一个最大限度地减

少学生的预期损失从方程 (1)。一个简单的选择下一个教学图像的策略是随机选择的从数据集 D 的随机抽样样本 (SRND) 做没有模型的学生, 因此无法适应他们的能力。这种缺乏适应可以在方法 1) 重复呈现教学实例学生已经学会的概念, 和 2) 没有直接加强的概念, 学生有显示自己 (通过反馈) 是不确定的。

杜和灵 [14] 提出了一种称为“最坏预测”的策略, 在这里, SWP, 是不确定性采样相关用于主动学习 [38]。然而, 与在主动学习, 在机器教学, 计算机是否有访问地面真理标签。他们的策略选择下一个教学图像作为预测偏离大多数的地面真理, $X_T = \operatorname{argmin}_X \hat{P}(Y | x)$, (2) 在 $\bar{y} = \operatorname{argmax}_y P(Y | x)$ 是真实的类标签老师知道。这种方法的缺点是这很容易提出的教学图像, 作为异常值在当前模式下, 他们往往是高度不确定的。这个问题的一个可能的解决办法是重量数据通过某种局部密度的测量空间例如 [38, 15]

3.1.1 预期减少错误的教学

我们的教学策略, 我们称之为先知, 灵感从最佳的抽样方法发现在学习活动 [36, 45, 29]。不像 SWP, 先见选择教学如果正确的话, 图像将有最大的减少未来错误的图像, 是不在教学中, 杜 = $D \cap DT$, 哪里 $X_T = \operatorname{argmin}_X P(\bar{y} | x)$, $\bar{y} \in \text{杜} (1 - \hat{P}(X_T, \bar{Y}_P) | \bar{y} | \text{西})$ 。(3)

在这里, $\hat{P}(X_T, \bar{Y}_P)$ 我是学生的最新估计条件分布如果他们显示 X_P 反过来正确地贴标签。这一战略有利的它首先集中在高密度区域的属性在特征空间中, 并为学生的提高, 提炼这些区域之间的边界。在语境中主动学习, 这被称为探索与开发贸易。这与方法有关学习所倡导的课程学习, 其重点首先, 简单的概念, 并逐步增加的难度 [5]。

3.2. 学生建模

在这项工作中, 我们近似学生的条件鉴于教学集分布, $\hat{P}(Y | X, DT)$, 利用图基于半监督学习 [41, 44]。使用高斯随机场 (GRF) 的半监督方法 [44], 我们可以宣传学生的班级标签的估计对于当前的教学设置, DT , 未观测到的图像杜通过定义一个相似度矩阵 $W \in \mathbb{R}^{N \times n}$. 效益使用基于图形的的方法是, 我们不需要直接在特征空间中工作, 并可以使用我们, 图像对之间的相似性。这给我们的灵活性允许使用特征向量定义相似从图像中提取, 人提供的属性, 或使用距离度量学习 [23]。

如果我们给出了一个特征表示为我们的教学集, 相似性计算我们的一个常用方法两个图像之间采用 RBF 核函数我们 = $\exp(-\gamma \|x - x_j\|^2)$ (4) 在这里, 是一个长度尺度参数, 控制多少相邻图像相互影响。使用矩阵表示法对 [44], 我们定义了 $n \times C$ 矩阵 $F = \hat{P}(Y | X, DT)$, 其中每个元素的 $F_{ij} = \hat{P}(Y = C_j | X_i)$ 。我们可以繁殖从以学生为教学提供的标签信息, 编为 $|DT| \times C$ 矩阵 FT , 到未标记图像赋, 福 = $(S - W) \cdot FT$, (5) 其中 S 是一个条目 $S_{ii} = 1$ 我们对角矩阵。所有在英国金融时报的条目是 0, 除了人类学习者估计 (正确或不正确) 教学的类别标签图像西, 我们设定的 $F_{ij} = 1$ 。五是相似的为不可观测的图像矩阵的全矩阵的一个子集在 [44], 我们可以有效地评估方程 (3) 使用标准的矩阵运算来进行数据集 2000 像在一秒钟内使用未经优化的 Python 代码。

4. 实验

为了验证我们提出的多类教学策略, 我们进行了真正的人类受试者的研究。参与者通过招聘机械特克 [1], 和相互作用使用我们的系统远程使用我们的自定义网络接口, 基于 Python 的 Django Web 框架的建立 [2]。

4.1. 数据

对于我们的实验中，我们选择了四个不同的数据集，总结在表 1。确保教学任务对参与者和一个射击学习是有挑战性的不可能的，我们选择了小的类间变化的数据集大类内变异。例如图像每个类都在图 2 中给出。不像标准分类数据集的日常对象例如，[16, 19]，我们的数据集包含图像类别是具有挑战性的非域专家之间的歧视，因为他们是由罕见的类组成。两个数据集，“蝴蝶”和“海底”，进行整理由本文作者从正在进行的科学视觉识别的研究。“蝴蝶”是一种一个较大的集合的英国蝴蝶图片集的子集一个博物馆收藏了 100 年的时间。海底是一组水下物种的图像一项研究试图衡量拖网在水下的影响生物多样性。这两个数据集是策划和注释由领域专家。

表 1: 使用的数据集的汇总，显示的数字类的最小和最大数量每级图像。图 4: 在我们的实验中使用的 2 个数据集的示例图像。每节课三个随机图像。值得注意的是，这些图像是具有挑战性的分类，因为它们表现出大量的类内变化。此外，'海底'的图像是特别困难的，因为它们被捕获的“在野外”，并包含闭塞和杂波。

使用公共提取的图像特征[21]的 ConvNet。对于每一个数据集，我们计算功能，使用网络预训练的 ImageNet2012 挑战数据集[37]。然后我们微调了完全连接层使用已知的地面真理类

我们的数据集，它产生了一个单独的标签每个数据集的 ConvNet。构造相似矩阵 $W(4)$ ，我们降低了 ConvNet 维度特点从 4096 到 50，使用主成分分析，并设置长度所有数据集的比例参数为 0.025。在我们最初的实验中，我们探讨了自定义设计的猪为基础的功能 [10]，我们发现表现更糟的相比我们微调 ConvNet。在这里，额外的监督信息在微调过程中产生一个表示在同一个类的图像更顺利分布在特征空间。

一个更好的特征空间与学生的相似性的观点应该受益所有的概率策略同样。这也可能通过众包计算之间的教学图像的相似性从一组用户的图像排名，例如[32]。不过，我们发现我们的 ConvNet 特征是一个很好的平衡减少额外的督信息量每一个教学任务和学生的要求性能。我们的项目可以使用代码和数据网站。

4.2. 实验设计

为了评估我们的教学算法，我们进行了实验在受试者通过机械土耳其 [1]。此前，克伦普等人。[9]已证明可以复制经典类学习的结果利用机械特克实验。使用类似实验设置为 [31]，我们的参与者首先提出有了一系列的教学图像，然后其次是一个序列的测试图像。对于每一个实验中，参与者首先告诉他们被要求学习许多课程。教学开始使用交互式教学环化的方向图 1。对于每一个教学形象，参与者都是首先显示图像，要求估计其类标签在我们的网页界面上点击按钮，然后提供正确的答案。接收参与者的估计类标签，教学策略更新它的学生模型和选择下一个图像显示。与教学相比较，没有纠正反馈的形式真正的类标签在测试阶段提供。测试轮仅用于评估目的，是没有必要的真实教学情境。测试图像被随机选择对于每一个参与者，与每一个平等的，和被排除在可能的教学。每一个参与者都有一个随机数据集

从表 1，结合随机教学策略。对于每一个数据集，示出的教学图像的数量设置为三倍的类，和十倍的测试数。这样，教学的长度测试轮与复杂度成正比任务。我们进行了较长的教学试验 (>40 图像)和测试之间定期的教学实现学习曲线的图像。然而，我们发现通过反馈，学生变得无聊和沮丧随着强迫的延迟，鼓励他们放弃。值得注意的是，我们的教学任务是显著的更

难比大多数人群来源的图像注释任务。在典型的注释任务，工人已经拥有很强的先验知识的概念，而在我们的教学任务中，参与者不可能有优先领域专业知识。我们调查的参与者在开始的任务，以确保他们没有事先的任务知识我们拒绝了那些声称拥有的结果即使是中等程度的类。因此，学生对第一个教学形象的回答总是随机猜测。为了避免看似按随机的，我们也拒绝了这些结果图像的平均响应时间太快（3 秒）在测试过程中。鼓励认真学习，我们支付工人的奖金，如果他们得分较高在测试过程中的阈值。在丢弃噪声的参与者，我们收集的结果从 25 和 35 参与者每一策略/数据集组合。

4.3. 基线策略

除了概述的基本教学策略 3.1 节，我们也比其他两个基线 SCC 和 sbatch。SCC，或类的质心，在计算为每个类的一个给定的数据集的特征空间的质心，和学生只提出的图像表示这些中心在教学。教学形象随机选择从其中的一个质心。如果小的类内差异，如果一次性学习是有可能的，或者说如果课程是熟悉的对于学生，我们希望这个基准来执行很好。最后的底线，类似于离线批量教学算法如 [39]。这里，订购离线计算的教学图像。我们计算用 SEER 算法排序，但假设如果显示了一个形象，学生将永远贴上标签正确地。鉴于这个假设，教学的选择图像是确定的，是相同的所有学生不管他们的反应。离线二进制的最近策略教学，如 [39]，不直接适用比较，因为我们在具有挑战性的互动多类分类场景。

4.4. 人体实验

从人的参与者的结果总结在表 2。个人数据集的结果被描述在图 3，其中测试图像的平均数回答正确地显示为每个

数据集和策略组合。我们可以看到，我们的先知方法优于“中国”，“蝴蝶”，和其他教学策略“海底”数据集。在这三个，我们的方法是一致的最佳的执行，而其他的方法有不同的性能根据特定的数据集。正如我们所看到的从表 2 中，没有明确的“秒最佳”的方法，和离线 sbatch SWP 策略往往与不确定性优于随机 Sr-Nd。先知的表现最为明显的“海底”数据集，其中也包含最随意的图像，由于数据的采集在野外拍摄的相机，而不是整齐地成像在受控的实验室条件下。在测试过程中的平均时间为不同的策略，计算时间之间被显示的测试图像并提交一个答案，在表 2。使用我们的方法的参与者往往会回答更多快速与其他策略相比。Sr-Nd 和 SCC 也具有较低的响应时间，但学生较差的性能测试时间可能表明一个虚假的信心水平。表 3 提供的统计 significancetable 2 p 值：平均响应时间测试中的参与者，和测试集的分数的，在所有的数据集。我们的结果。双尾试验进行了一个空假设我们的方法的分数分布在所有的数据集，和竞争的方法，是统计类似的，基于高斯假设。p 值得到的标准测量范围内的 0.05 测试的统计意义，表明我们的结果是不因偶然。图五显示了 4 个平均学习曲线教学中的教学策略。平均得分为每 10% 个进度间隔（通过培训集）计算的平均数，正确的反应在这一点上所有的学生和数据集教学阶段。请注意，这是不等同于真正的学习曲线，由于图像被选为主动教学生，而不是评估他们的表现快照。我们看到了提高认识的总趋势进一步教学图像的速度。然而，鳞状细胞癌给出了一个假意义上的表现，因为相同的质心图像反复证明，因此学生 overfits 这些图像，通常在测试过程中不能概括。不像其他的，基于不确定性的 SWP 的策略比较平坦的学习曲线，因为异常图像 shownare 具有挑战性的学习。

这欠给学生对每一类变异的理理解只有一个弱。图 5 显示的教学图像的例子对学生的五个策略与‘中国’数据集。我们看到先知适应错误的的能力响应，注意的“干”类的原因要一个不正确的前一个答案，在返回教‘草’由于其先前不正确的答案，终于探索学生对“丘”的认识。在另一方面，sbatch 无法适应其教学和研究尽管学生的“丘”和“干”的教学我们发现，人类学习者通常假定在教学过程中，往往会集中单峰分布一个单一的物种在整个属。

5. 结论

机器教学有可能使人类学习理念，无需人对人的专家辅导。通过自动调整课程给学生的

能力和性能，教学可以进行在它是困难或非常昂贵的情况从一个人获得直接进入领域的专业知识教师。在这项工作中，我们已经朝这个方向迈出了一步提出了一种交互式多课堂教学策略。它的目的是为了给学生呈现的教学图像将最翔实，给出了他们的在线估计目前知识。不像其他建议的策略，我们

不太可能教野值，并因此，不要浪费时间不具有代表性的图像显示。类似课程学习[5]，我们的战略最初侧重于代表图像，然后介绍了更困难的时间，作为学生的表现提高。

5.1. 今后的工作

目前，我们提出的教学图像的学生一个在一个时间。在未来，我们计划调查不同的图像显示方法。如成对的可视化比较[22]，并突出局部区域[11]或部分[6]，可证明更有效地在输送歧视不同类别的细节和特点。一些图像本质上比其他的更令人难忘[12, 20]，并将这些措施纳入教学图像选择也可以提高测试时间性能。在课程学习中，任务难度随着性能的提高而增加提高。在未来的工作中，我们也

将探讨其他的教学范式，如螺旋法教学[7]。在螺旋式学习中，介绍了新的分类随着时间的推移而不断地重新强调早期概念，以确保他们成为致力于记忆。

鉴于我们现在可以教人类视觉分类自动化的任务，在未来的工作中，我们打算调查我们可以提取什么额外的信息从我们的学后的教学。与…对比机器，研究表明，人类可以学习与理想化版本的数据，可以有不同的分布从测试集[17]。探究式教学适应问题可以让我们获得注释对于数据，这是非常不同的，从我们的教学。最后，我们假设我们的特征空间是相关的与学生的相似概念。可能更有效共同估计学生的当前能力和教学的相似性概念。

References

- [1] Amazon Mechanical Turk. <https://www.mturk.com>, 2014.
- [2] Django Web Framework. <https://www.djangoproject.com>, 2014.
- [3] F. J. Balbach and T. Zeugmann. Recent developments in algorithmic teaching. In *Language and Automata Theory and Applications*. 2009.
- [4] S. Basu and J. Christensen. Teaching classification boundaries to humans. In *AAAI*, 2013.
- [5] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. In *ICML*, 2009.
- [6] T. Berg and P. N. Belhumeur. How do you tell a blackbird from a crow? In *ICCV*, 2013.
- [7] J. S. Bruner. *The Process of Education*. Harvard University

- Press, 1960.
- [8] M. Cakmak and M. Lopes. Algorithmic and human teaching of sequential decision tasks. In AAAI, 2012.
- [9] M. J. Crump, J. V. McDonnell, and T. M. Gureckis. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PloS one, 2013.
- [10] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005.
- [11] J. Deng, J. Krause, and L. Fei-Fei. Fine-grained crowdsourcing for fine-grained recognition. In CVPR, 2013.
- [12] A. Deza and D. Parikh. Understanding image virality. In CVPR, 2015.
- [13] T. Doliwa, H. U. Simon, and S. Zilles. Recursive teaching dimension, learning complexity, and maximum classes. In Algorithmic Learning Theory, 2010.
- [14] J. Du and C. X. Ling. Active teaching for inductive learners. In SDM, 2011.
- [15] S. Ebert, M. Fritz, and B. Schiele. RALF: A Reinforced Active Learning Formulation for Object Class Recognition. In CVPR, 2012.
- [16] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes (VOC) challenge. IJCV, 2010.
- [17] G. Giguère and B. C. Love. Limits in decision making arise from limits in memory retrieval. PNAS, 110(19):7613 - 7618, 2013.
- [18] S. A. Goldman and M. J. Kearns. On the complexity of teaching. Journal of Computer and System Sciences, 1992.
- [19] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. 2007.
- [20] P. Isola, J. Xiao, D. Parikh, A. Torralba, and A. Oliva. What makes a photograph memorable? PAMI, 2013.
- [21] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. arXiv:1408.5093, 2014.
- [22] A. J. Joshi, F. Porikli, and N. Papanikolopoulos. Breaking the interactive bottleneck in multi-class classification with active selection and binary feedback. In CVPR, 2010.
- [23] B. Kulis. Metric Learning: A Survey. Foundations & Trends in Machine Learning, 2012.
- [24] N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. Soares. Leafsnap: A computer vision system for automatic plant species identification. In ECCV, 2012.
- [25] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollr, and C. Zitnick. Microsoft COCO: Common

- Objects in Context. In ECCV, 2014.
- [26] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang. Casia online and offline chinese handwriting databases. ICDAR, 2011.
- [27] C. Long, G. Hua, and A. Kapoor. Active visual recognition with expertise estimation in crowdsourcing. In ICCV, 2013.
- [28] B. C. Love. Categorization. In Oxford Handbook of Cognitive Neuroscience, pages 342 - 358. 2013.
- [29] O. Mac Aodha, N. D. Campbell, J. Kautz, and G. J. Brostow. Hierarchical Subquery Evaluation for Active Learning on a Graph. In CVPR, 2014.
- [30] O. Mac Aodha, V. Stathopoulos, G. J. Brostow, M. Terry, M. Girolami, and K. E. Jones. Putting the scientist in the loop - accelerating scientific progress with interactive machine learning. In ICPR, 2014.
- [31] D. L. Medin and M. M. Schaffer. Context theory of classification learning. Psychological review, 85(3):207, 1978.
- [32] P. O' Donovan, J. L. Leeks, A. Agarwala, and A. Hertzmann. Exploratory font selection using crowdsourced attributes. ACM Transactions on Graphics (TOG), 2014.
- [33] K. R. Patil, X. Zhu, Ł. Kopeć, and B. C. Love. Optimal teaching for limited-capacity human learners. In NIPS, 2014.
- [34] V. C. Raykar, S. Yu, L. H. Zhao, A. Jerebko, C. Florin, G. H. Valadez, L. Bogoni, and L. Moy. Supervised learning from multiple experts: whom to trust when everyone lies a bit. In ICML, 2009.
- [35] J. J. Richler and T. J. Palmeri. Visual category learning. Wiley Interdisciplinary Reviews: Cognitive Science, 2014.
- [36] N. Roy and A. McCallum. Toward optimal active learning through sampling estimation of error reduction. In ICML, 2001.
- [37] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge. arXiv:1409.0575, 2014.
- [38] B. Settles. Active Learning. Morgan & Claypool, 2012.
- [39] A. Singla, I. Bogunovic, G. Bartók, A. Karbasi, and A. Krause. Near-optimally teaching the crowd to classify. In ICML, 2014.
- [40] P. Welinder, S. Branson, S. Belongie, and P. Perona. The multidimensional wisdom of crowds. In NIPS, 2010.
- [41] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In NIPS, 2004.
- [42] X. Zhu. Machine teaching for bayesian learners in the exponential family. In NIPS, 2013.
- [43] X. Zhu. Machine teaching: An inverse problem to machine learning and an approach toward optimal education. AAAI Conference on Artificial

Intelligence (Senior Member Track),
2015.

[44] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In ICML, 2003.

[45] X. Zhu, J. Lafferty, and Z. Ghahramani. Combining active learning and semi-supervised learning using Gaussian fields and harmonic functions. In ICML workshops, 2003.

[46] S. Zilles, S. Lange, R. Holte, and M. Zinkevich. Models of cooperative teaching and learning. JMLR, 2011.

2624

