

指导教师： 杨涛

提交时间： 2016/3/20

# CVPR2015 Paper Translation

No: 01

姓名： 姚起程

学号： 2013302586

班号： 10011305



## 理解图像的病毒式传播

Arturo Deza  
UC Santa Barbara  
deza@dyns.ucsb.edu

Devi Parikh  
Virginia Tech  
parikh@vt.edu

### 摘要:

社交网站上的在线内容的病毒式传播是一种重要并且深奥的现象，往往在各个领域被研究，如营销学，心理学和数据挖掘。在本文中，我们从计算机视觉的角度研究病毒的图像。我们引进 reddit 的三个新的图像数据集，并使用元数据书签交易定义一个病毒式传播的分数。我们培养与国家的最先进的图像特征分类预先字典的单个图像，在图像对相关病毒式传播和病毒图像的主导话题的病毒式传播。我们也把人的表现与机器的表现作比较。我们发现，电脑在低级别的功能表现不佳，而高层次的信息是预测病毒式传播的关键。我们通过编码属性相关的语义信息。我们确定了 5 个关键与病毒式传播相关的属性。我们创建一个基于图片属性的特征描述，可以预测相对病毒式传播并且有 68.10% 的准确度 (SVM + 深相对属性)

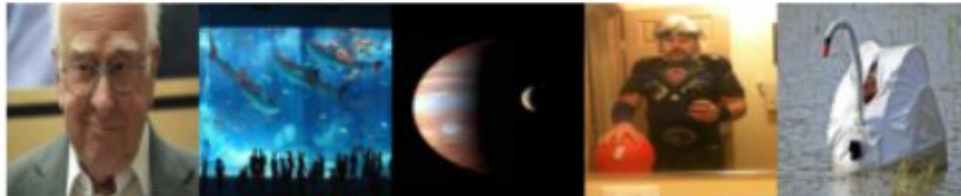
比人类的 60.12% 更好，图像的基于属性的表征。最后，我们研究图像网路效应的人预测如何与其中图像被看作不同的“CON-文本”，如相邻图像的影响，最近观看的图像，以及该图像标题或标题不同而不同。这项工作是在理解复杂但重要的图像扩散现象的第一步。我们的数据集和注释将被公布于众。

### 简介:

我应该使用什么样的图片来开始使得我们比 Instagram 的更加抢眼？哪个图片标题将帮助传播一种代表性不足的令人震惊的消息？如果我想以百万计的观看，我应该把猫的图片放在我的 YouTube 视频，？这些问题困扰着专业人士和日常的互联网用户。广告，营销策略，政治运动，非营利组织，社会事业，作家和摄影师，仅举几例，取决于他们的能力所能达到的并被注意



(a) Example viral images.



(b) Example non-viral images.

图 1: 上图: 在我们的数据集高病毒成绩图像描绘互联网“名人”模因。“脾气暴躁的猫”; 下图: 在我们的数据集低病毒成绩图像。彼得·希格斯的图片(希格斯粒子)很流行, 但没有转贴多次, 因此是不被视为病毒。通过大量的人, 了解是什么使病毒的内容也因此被广泛地被市场研究人员研究[7, 4, 11, 5]。

许多因素, 如当图像被上传的星期和一天中的时间, 与图像等所使用的标题影响图像是否去病毒与否[25]。到什么程度是病毒式传播要依赖这些外部因素, 以及病毒式传播的多少取决于图像内容本身? 如何才能国家的最先进的计算机视觉图像特征和人类预测病毒式传播? 它的视觉属性与图像相关病毒式传播?

在本文中, 我们来解决这些问题。我们引进从 Reddit 收集的三个数据集和一个扩散分数。我们的工作为更深的研究标识出几个有趣的方向, 其中计算机视觉技术可以被用来承担理解和预测图像病毒式传播这个复杂的问题。

相关工作:

大多数现有的工作[26, 2, 30]研究人们如何在它被发布后在社交网络上分享内容。他们在内容发布后不久使用网络动态内容来检测迎面而来的滚雪球效应, 预测内容是否会去病毒或不会。我们认为, 在内容被公布后预测病毒式传播是在某些应用中为时已

晚。对于图形的设计者而言, 去尝试各种各样的设计来查看是否它们成为病毒或没有是不可行的。在本文中, 我们有兴趣了解内容本身(即使在他被发布到网上之前)和其病毒化的可能性。

存在有可能去病毒的种类的内容的几个定性理论[4, 5]。只有少数作品定量分析的内容, 比如推文[32]和纽约时报的文章[6]预测其病毒式传播。然而, 尽管它们是很大一部分我们的在线体验, 视觉媒体内容和他们的病毒式传播之间的联系一直没有进行分析。这形成了我们工作的重点。

文本数据如微博的网路效应进行了研究[27, 32]。扩散性能被认为是依赖于它们的内容和类似的嵌入 URL 的和标签功能。通常, 在网络上的内容的扩散已经研究了比的原因[30]以上。Leskovec 等人的工作。[26]车型在通过随机模型个人组成的网络传播的建议, 而布泰尔等。[8]接近病毒扩散为流行病学的问题。

是什么使人们共享内容定性理论已经在市场研究被提出。伯格等人。[4, 6, 5]例如假设一组 STEPPS 那意味着社会的货币, 触发器, 易于情绪, 公

众（宣传），实用价值和故事让人们分享。

分析病毒的图像已收到很少的注意。Guerini 等[18]提供了一个非匿名的社交网络上（Google）对情绪和网络效应[17]之间的联系的低级视觉数据和普及之间的相关性，以及。斯拉等人。[23]最近研究了图像的普及作为视图照片有 Flickr 上的数量来衡量。自然的照片然而，无论以前的作品[18, 23]只提取图像统计（Google+的，Flickr 的）。图像和 Reddit 上的社会交往有质的不同（例如许多 reddit 的

图像编辑）。从这个意义上讲，图像的质量是最相似的我们是 Wang 等人的同时引入病毒梅梅发生器，结合自然语言处理和计算机视觉（低级别的功能）[37]。然而，我们的工作再往进深的固有可视内容（如高级图像的属性），可视上下文周围的图像，颞 CONTEX 和文本上下文中图像 virality. Lakkaraju 等的作用。[25]分析了一天，一周中的天，重新提交，标题，类别，等数的时间的的影响。上 Reddit 上的图像的病毒式传播。然而，它们不分析图像本身的内容。



图 2: 病毒式传播 (V 高) 与图像中的普及 (一小时)。所有图像具有相似的人气比分，但他们的病毒式传播的分数有很大不同。“脾气暴躁的猫”是比伯希格斯多种病毒由于重新提交的号 (m h) 中，可以播放在我们的网路效应度量 V h 的关键作用。显然，病毒式传播和普及是两个不同的概念。

在计算机视觉几个作品已经研究复杂的元现象（相对于理解图像中的“文字”等内容作为对象，场景，3D 布局设计等）。伊索拉等。[20]发现，一些图像始终更令人难忘比别人跨学科分析图像内容，使图像难忘的[19]。图像美学在[14]，图象情感了研究中[10]，和[12]中的对象技术的认可。对象[31]，属性[36]以及由该人第一次提到它们中的图像的描述的可能性中定义的场景[3]的重要性也进行了研究。我们研究图像病毒式传播的独特复杂的现象。

### 三. 数据集和实地扩散

#### 3.1 扩散分数

reddit 的是世界各地的病毒含量的主要引擎。上个月，它有代表的每一个

国家超过 170M 独立访问者。它有一个巨大的各种主题的超过 353K 类别 (subreddits)。我们专注于图像内容。这些图像有时候是珍贵的照片或者描绘滑稽的照片，荒诞的情况下，或 Redditors 通过图片分享个人情绪化的时刻，或通过图像表达他们的政治或社会的意见等等。每个图像可以赞或由用户反对。病毒性的内容往往被重新提交多次，因为它跨用户 3 的网络传播。病毒的图像因此具有许多赞，很少反对的那些，和已被经常重新提交的不同用户。后者是从什么区别流行病毒式传播。先前，Guerini 等。定义了多个病毒式传播的指标作为赞，分享或评论，科斯拉等。定义流行观看次数和 Lakkaraju 等。定义受欢迎程度赞的数量。我们发现，由赞和病毒式传播的数量也占补交（接下

来详细定义) 定义流行之间的相关性为-0.02。这表明定量这两种现象之间的区别。参照图 2 定性的例子。的焦点

本文是研究图像病毒式传播 (而不是普及)。

让得分  $S$  是 upvotes 的数目之间的差和 downvotes 在其  $N$ -收到第重新提交图像  $h$  至一个类别。设  $t$  是图像的重新提交的时间,  $c$  是类别 (版 (Subreddit)), 以它被提交。  $S$  是所有提交的  $C$  类别在时间  $t$  的平均成绩。我们定义了一个要的比率在重新提交  $n$  至所有图像的平均得分像  $h$  的成绩张贴在小时 [25] 的类别。

$$A_h^n = \frac{S_h^n}{S_c^t}$$

我们加一个偏移量  $S$  以便最小得分分  $\hat{h}$  分钟  $n_s$  是 0。我们定义的整体 (所有类别) 病毒式传播得分图像  $H$  作为

$$V_h = \max_n A_h^n \log \left( \frac{m_h}{\bar{m}} \right)$$

其中,  $m$  是被重新提交的倍图像小时数, 和  $\bar{m}$  是平均次数任何图象已被重新提交。如果图像往往重新提交, 其病毒式传播的分数就高。这确保了, 成为流行的时候被贴, 但没有转贴图像, 不被认为是病毒 (图 2)。这些往往涉及图像, 其中内容本身不太相关, 但目前的事件引起人们对图像, 如最近的悲剧, 新闻快讯, 或个人的成功

故事例如 “妈呀, 我减掉了 40 磅, 在 2 周内”。在另一方面, 具有多种图像提交似乎大约多种情况不同职称的 “弹性”, 可以说是, 本质上病毒。实例示于图图 1 (a)。

### 3.2 病毒图像数据集

我们用来自 riddit 数据收集的图片来自 [25] 创造我们的数据集。Lakkaraju 等。 [25] 度过一个为期 4 年的 reddit 的 132K 条目。条目通常对应于同一图像的多个提交。我们只包括从类别 (subreddits) 即至少有 100 提交我们的数据图像, 使我们有  $m$  的公式 2, 我们丢弃的 GIF 动画的准确测量。这给我们留下了 20 类共 10078 的图像, 每个图像  $\bar{m} = 6.7$  意见书。我们决定从 reddit 的, 而不是其他社交网站如 Facebook 和 Google+ [18], 因为用户在 Reddit 上 “4THELULZ” 发表图片用图像 (即只是为了好玩), 而不是个人的社会知名度 [6]。我们也更喜欢使用 reddit 的, 而不是 Flickr 的 [23] 因为在 reddit 的图像被匿名发布, 因此, 它们的产生最纯净 “互联网拖钓” 的形式。

### 3.3 病毒和非病毒图像数据集

接下来, 我们在研究这种复杂创建包含 250 个最多和最少的病毒的图像每个使用等式 2。网路效应比分两组图像中的这种鲜明的对比为我们提供了干净的二分法探索作为第一步骤 500 的图像数据集 现象。回想一下, 非病毒的图像包括这并没有得到足够的赞的图像, 以及那些可能已经在一个多提交赞, 但没有公布多次。



图 3: 示例图片来自 3 个最病毒类别 (从上到下): 有趣, 跆拳道, AWW。

### 3.3.1 随机对数据集

与在上面的数据所表示的清洁二分法相反, 我们还创建图像对的一个数据集, 其中在一对中的两个图像的病毒式传播的差少明显。我们从 >10K 的图像从配对使用随机图像 250 最广为流传的图像随机图像与网路效应比中值网路效应低。同样, 我们从配对使用随机图像 250 至少病毒图像的随机图像比位病毒式传播高。我们收集 500 个这样的对。去除这种情况发生, 从顶部/底部 250 病毒图像有两个图像对给我们留下了 489 对。我们报告此数据集提供了最终的人力和计算机的结果, 和在表把它称为 (500 p) 的 2. 培训于其他 4550 对可以从剩余 10K 的图像通过配对与上述位数病毒图像形成完成低于中位数病毒图像。

### 3.4 病毒分类数据集

搞笑, 跆拳道, 噢, 无神论和游戏: 对于我们最后的数据集, 我们与五个最病毒类工作。我们确定, 只有在类别, 而不是其他的一个是病毒的图像。要做到这一点, 我们计算图像的病毒式传播的分数之间的比例相对于该给它的最高分, 它被提交到所有类别中的类别, 类别, 给了它第二个最高分。

$$V_h^c = \frac{V_h^{c^1}}{V_h^{c^2}}$$

那是

其中  $V$  是  $C$  类, 给了它第  $k$  个所有类别中得分最高收到的病毒式传播比图像小时。

$$V_h^{c^k} = A_h^{c^k} \pi \left( \log \left( \frac{m_h^{c^k}}{\bar{m}_h} \right) \right)$$

其中  $A$  是如等式 1 所定义的, 给了认为图像  $h$  的提交的所有类别中的第  $k$  个最高得分的类别,  $\pi(x)$  是  $x$  的百分等级, 是次像  $h$  的数量被提交给这一类, 而  $MH$  是时代形象小时平均数提交到所有类别。我们取百分等级, 而不是实际的日志值, 以避免在等式 3 中的比率负值。

要形成我们的数据, 我们只考虑顶部 5000 在我们的病毒图像数据集(第 3.2 节) 排名病毒的图像。这些包含 1809 好笑, 522 跆拳道, 234 噢, 123 无神论和 95 的游戏图像。其中, 我们选择的是每过公式 3 中的最高分, 形成我们的病毒分类数据集类 85 的图像。

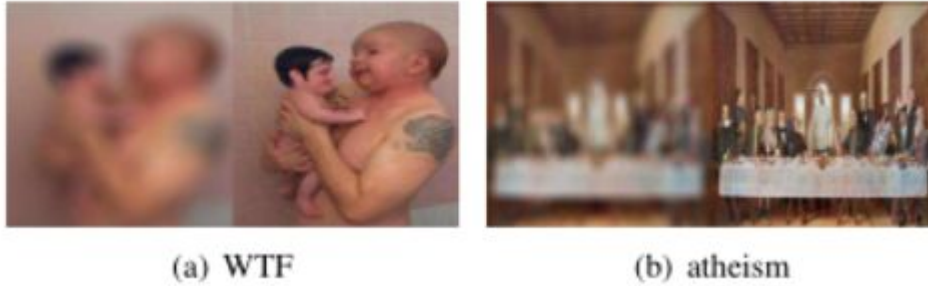


图 4: 通过病毒的图像模糊时间语境启动的例子。看着左边两个 (一) 和 (b) 图像, 你认为什么实际的图像描绘? 难道你的图像的期望变成是准确的?

#### 四. 理解图片的扩散

考虑图 4 的病毒的图像。凡脸上交换 [9], 上下文启动 [33] 和场景要点 [28] 使图片来自我们可能期望在乍看之下完全不同。在 NLP 研究类似的情景理解的语义 “这是她说了什么!”

笑话 [24]。我们推测, 也许没有出现这样的视觉挑战或矛盾的图像 - 在图像的语义感知不上的图像仔细检查显著改变 - 是 “无聊” [26, 6] 并不太可能是病毒。这个矛盾不必从图像中的对象或属性茎, 还可以从图像的上下文中上升: 是它周围的图像的图像, 或图像之前观察的图像, 或图像的标题等。或许这些不同的背景和图像的所得解释不一致之间的相互作用是必要的, 以模拟视觉双关语导致图像网路效应。考虑到这一点, 我们定义了四种形式的背景下, 我们将研究探索影像病毒式传播的。

1. 固有上下文: 这是指可视内容是固有的图像的像素。
2. 近处上下文: 这是指包围图像 (空间附近) 的图像的视觉内容。
3. 时空上下文: 这是指图像的视觉内容之前的图像 (时间附近) 看到。
4. 文本语境: 这种非可视上下文是指图像的标题或标题。这些头衔有时会表现为视觉内容 (例如, 如果它是 Photoshop 处理)。一个字涂鸦文字和内在环境, 并需要 NLP 与计算机视觉的理解。

#### 4.1 内在环境

我们首先检查人类和机器是否能够仅仅通过在图像预测, 无论是病毒图像或没有, 和什么该图像的占主导地位的主题 (最合适的类别) 的。对于机器的实验中, 我们使用使用实施国家的最先进的图像功能, 如无咖啡因, 深厚的功能 [15], 依据 [28], HOG [13], 微小的图像 [35] 等 [38]。我们开展亚马逊的 Mechanical Turk (AMT) 人体研究。我们怀疑, 熟悉 reddit 的工人可能在识别病毒式传播和类别比那些不熟悉 reddit 的不同的表现。所以我们创建了一个资格考试, 每个工人必须做任何我们的任务之前采取。测试包括约广泛传播 reddit 的模式和行话, 这样任何人都熟悉 reddit 的可以很容易地得到高分, 但工人谁不会得到一个非常贫穷的分数问题。我们阈值处理这个分数, 以确定熟悉 reddit 的与否工人。每个任务是由 20 名工人完成。图像是在  $360 \times 360$  所示。机床的精度进行了计算在相同的测试设定为人类的研究。人力精度使用跨工人多数表决来计算。结果报道为工人的不同子集 (1) 的精度 (例如, 那些熟悉 Reddit 和那些不) 可各自比整体精度低, 和 (2), 我们不能在我们的结果报告的误差棒。我们发现, 精度最高的工人穿过我们的任务由  $\pm 2.6\%$  变化。平均来说, 工人响应的 73% 的匹配每个图像的多数表决响应。

##### 4.1.1 预测话题

我们开始我们的话题分类实验，在实际应用是帮助用户决定提交自己的形象哪一类。我们用我们的病毒分类数据集（3.4 节）。参照图 3. 图像一般做似乎从一类到另一个不同的。例如，属于 AWW 类别的图像似乎包含在图像的中心可爱的小动物，在无神论的图像似乎有文字或宗教符号，在 WTF 图像往往是明确的，容易激起人们的厌恶，恐惧的感觉和惊喜。

培训合格的 20 工人每个类别 55 图像样本蒙太奇后，他们取得了 87.84% 的类别识别精度 25 测试图像，大多数的混乱很有趣的游戏和图像之间。与书签交易事先熟悉没有影响，因为在训练阶段的准确度。使用各种功能的机器的性能可以在图中看到。图 5 (a)。

通过使用无咖啡因 6 [1]（机会精度将是 20%）中获得的 62.4% 的性能。机器和人的困惑矩阵可以在增刊上找到。

#### 4.1.2 预测病毒式传播

现在，我们考虑预测图像是否是病毒或不看它的内容，通过使用我们的病毒和非病毒的图像数据集（3.3 节）的更具挑战性的任务。

我们要求受试者在 AMT 他们是否认为一个给定的形象将是病毒（即“成为如 Facebook, Twitter, reddit 的, Imgur 社会网络工作的网站非常的病毒，等有很多人喜欢，再啁啾，共享或 upvoting 的形象呢？”）。分类准确率为 65.40%，其中的机会是 50%。

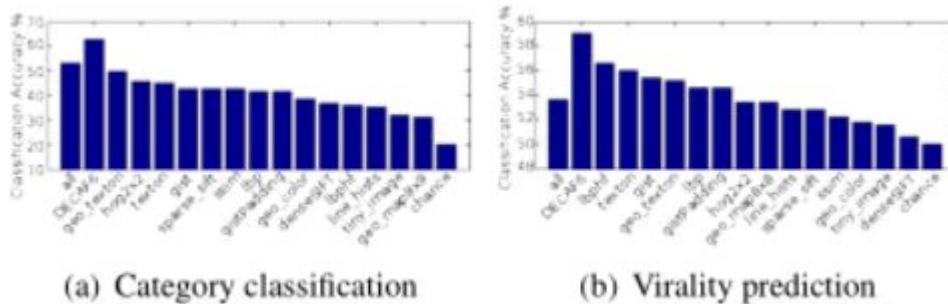


图 5: 在我们的病毒分类机械精度 (3.4 节)

和病毒与非病毒图像数据集（在顶部/底部 250 对 3.3 节测试），使用不同的图像特征。

在这些任务中，我们还问工作人员，如果他们以前见过的形象，基于与图像的熟悉拿到感为他们的偏见。我们发现，9%，1.5%和图像的 3% 已被 reddit 的工人，非 reddit 的工作人员和所有员工见过。虽然小样本集，该子集分类准确度均较高：75.27%，93.53%和 91.15%。需要注意的是病毒的图像有可能被非书签交易用户通过其他社交网络甚至看到。此外，我们发现，谁在一般熟悉 reddit 的工人有大约相同精度工人谁没有（分别为 63.24%和 63.08%）。然而，他们也有不同的分类策略。reddit 的工人们

40.64% 命中率，而非 reddit 的工人们 28.96% 命中率。这意味着，reddit 的工人更容易识别的图像为病毒，当他们看见一个人（但可能其他错误分类非病毒图像作为病毒）。非 reddit 的劳动者在调用图片病毒较为保守。低于 50% 都命中率表明朝向标签图像作为非病毒的一般倾向。这可能是由于现有超过在用于此实验的数据集中的病毒和非病毒的图像的不自然均匀的。总体而言，谁都没见过像之前不熟悉 reddit 的工人，可以预测优于偶然性，图像的病毒式传播。这表明，内在的图像内容是指病毒式传播，而像 reddit 的社区的形象病毒式传播不仅是一个滚雪球式的偶然煽动效应的结果。



使用我们的网路效应度量机的性能示于图。6. 其他指标可以在 supp.mat 找到。我们看到，当前视觉模型有这些病毒和非 viralimages 之间很难区分时，根据任何标准。支持向量机用

线性和对我们的~10k IM-年龄的数据集的 5 个随机分割非线性内核嚶使用每个类进行检测的 250, 500, 1000, 2000, 4000 的图像训练, 和 1039 的图像。

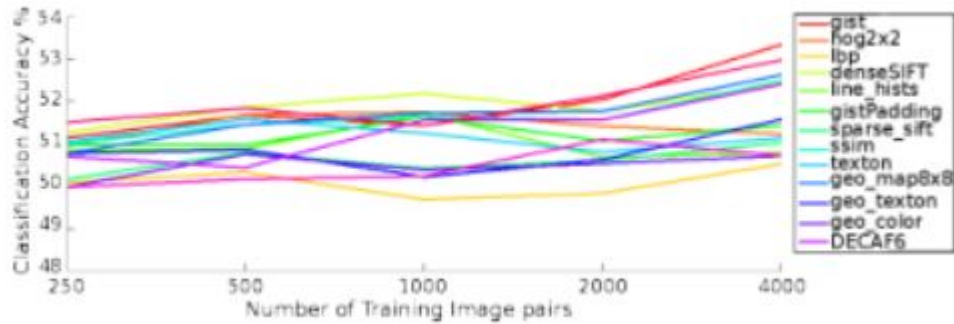


图 6: 用我们的病毒式传播公制机械精度横跨 5 个随机火车/测试分裂平均值, 测试集包含在每个时间 2078 随机图像。请注意, 所有描述符产生类似的结果的机会 (50%)。新颖的图像理解技术, 需要发展来预测病毒式传播。

机器上的同一组中使用多种特征来预测网路效应人类研究中所用的图像的性能示于图。图 5 (b)。于顶部和底部 2000 图像进行训练, 但不包括用于测试的顶部和底部 250 的图像。无咖啡因还提供 59% 达到最高的准确度; 这是上面的机会, 但比人的行为 (65.4%) 低。图像在 Reddit 上的很大的可变性 (整个纸张所示) 和国家的最先进的图像特征表现不佳表明图像病毒式传播的自动预测需要先进的图像理解技术。

#### 4.1.3 预测相对病毒式传播

预测单个图像的病毒式传播是人和机器都可一项艰巨的任务。因此, 我们考虑病毒式传播的相关预测。也就是说, 给定一对图像, 是更容易预测该两个图像的更可能是病毒? 在心理物理学, 这种设置是称为双替代强迫选择 (2AFC) 任务。

我们创建了一个由病毒随机图像, 从我们的病毒和非病毒的图像数据集 (3.3 节) 随机的非病毒图像的图像对。我们问工作人员这两个图像更有可能去病毒。精度是所有工人 4: 71.76%, reddit 的工人: 71.68%, 非 reddit 的工人: 68.68%, 在绝对明显的任务高于 65.40%, 而 50% 的机会。使用无咖啡因 6 图像特征的 SVM 得到了 61.60% 的准确度, 类似的绝对任务 SVM 分类准确度 (图 5 (b) 条)。

#### 4.1.4 相对属性和病毒式传播

现在, 我们已经建立了病毒式传播的非平凡部分不依赖于图像内容, 我们要了解什么类型的图像往往是病毒, 即什么样的图像特性与病毒式传播相关。我们对 AMT 科目注释同样对在上述实验中使用的图像, 相对属性注释 [29]。



个，增加网路效应预测之最。比如，有趣的图像往往是病毒，而且与动物的图像往往是病毒。但是，这是有趣的和有动物图像可能更容易被病毒。要添加的属性可以是属性本身（↑），或它的否定（↓）。这有助于处理那些负相关病毒式传播的属性。例如，合成产生是不美丽的图像更可能比要么合成图像的病毒生成或不美。通过这种方式，我们贪婪地添加属性。表 1 表明，协作与网路效应良好关联的属性。我们排除“可能去病毒”和“难忘”从这个分析，因为这些本身都是高层次的概念，并且不会增加我们病毒式传播的理解。

38 属性的组合导致了实现了 81.29% 的准确度一个网路效应预测。这可以被看作是网路效应的混合人机预测。的属性已被人类注释，但是属性已通过统计分析被选择。我们看到，这个显著优于单独人类（71.76%）和单独的机器（59.00%，见表 2）。一个可以训练该属性的预测，以进一步提升性能的顶部上的分类器，但是由表 1 提供的语义解释性将丢失。我们的分析

就开始给我们一个指示，其中图片属性需要可靠地预测自动预测病毒式传播。

我们还探索“属性吸”的效果：如果组合中的第一个属性是一个是负相关的病毒式传播相关，多么容易的是它从回收，使图像病毒？考虑这样的形象很“轻松”的场景（带负相关 viral-性）。是否有可能为图形设计人员通过改变图像的其他属性，使其病毒引发病毒式传播？图 7 (c) 示出的相关的轨迹随着更多的属性贪婪地加入到“种子”属性，该属性是正 (+)，负 (-)，或中性 (N) 的相关与病毒式传播。我们看到，在所有这些情况下，图像可以通过添加只是几个属性进行病毒。表 1 列出了哪些属性选择 3 种不同的“种子”的属性。有趣的是，虽然性是积极与网路效应相关，当动物接种，不性增加与网路效应的相关性。因此，当我们贪婪地选择我们的五个属性，相关性最好的病毒式传播的组合是：动物，合成产生的，不美的，明确的，而不是性感。

#### 4.1.5 自动化相对病毒式传播预测

	1	2	3	4	5
Attribute (+)	↑ synth. gen.	↑ animal	↓ beautiful	↑ explicit	↓ sexual
Virality Correlation	0.3036	0.3067	0.3813	0.3998	0.4236
Attribute (-)	↑ beautiful	↑ synth. gen.	↑ animal	↑ dynamic	↑ annoyed
Virality Correlation	-0.1510	0.2383	0.3747	0.3963	0.4097
Attribute (N)	↑ religious	↑ synth. gen.	↑ animal	↓ beautiful	↑ dynamic
Virality Correlation	0.0231	0.1875	0.3012	0.3644	0.3913

表 1: 具有病毒式传播人类标注属性组合的相关性。组合“引”第一属性。

Dataset	Classification Method	Performance
	Chance	50%
All images	SVM + images features	53.40%
Top/Bottom 250 viral (Section 3.3)	Human (500)	71.76%
	SVM + image features (500)	61.60%
	Human annotated Atts.-1 (500)	56.77%
	Human annotated Atts.-3 (500)	68.53%
	Human annotated Atts.-5 (500)	71.47%
	Human annotated Atts.-11 (500)	73.56%
	Human annotated Atts.-38 (500)	<b>81.29%</b>
Top/Bottom 250 viral paired with random imgs. (Section 3.3.1)	Khosla <i>et al.</i> Popularity API [23] (500 <sub>p</sub> )	51.12%
	SVM + image features (500 <sub>p</sub> )	58.49%
	Human (500 <sub>p</sub> )	60.12%
	Human annotated Atts.-5 (500 <sub>p</sub> )	65.18%
	SVM + Deep Attributes-5 (500 <sub>p</sub> )	<b>68.10%</b>

表 2: 在不同的数据集和方法相对病毒式传播的预测。

要创建一个自动化的相关病毒式传播的预测分类, 我们开始使用我们完整的~10k 图像数据集, 并有工人 AMT 做同样的任务, 第 4.1.4 节, 把它们分为病毒(上半部分在等级)与非病毒性(低半级), 并随机配对起来相对属性标注为前 5 执行从我们在图贪婪搜索属性。7 (三): 动物, 合成产生 (SynthaGen), 美丽的, 明确的和性。请注意, 我们所有的前五名属性是视觉。在混合人机网路效应预测组合属性为所有数据集的相关的轨迹可以在图中看到。图 7 (b)。

与所有的注释, 我们再通过 10 倍交叉验证训练相对属性的预测为每这些属性无咖啡因 6 深厚的功能 [15] 和 SVM 分类来获取所有的图像对 (第 3.3.1 节) 相对属性的预测。我们获得相对属性的预测精度是: 动物: 70.14%, Synthgen: 45.15%, 美: 56.26%, 显式: 47.15%, 性: 49.18% (可能: 33.33%), 由包括中性对。Futhermore, 我们得到的动物: 87.91%, Synthgen: 67.69%, 美: 81.73%, 显式: 65.23%, 性: 71.13% 为 +/- 相对的标签, 但不包括中性 (并列) 对 (可能: 50%)。结合这些自动属性预言 inturn (自动) 预测的病毒式传播, 我们得到了 68.10% 的准确度。如果我们用这 5 个属性的地面实况相对属性的注解, 我们达到 (65.18%) 的精度, 比直接从图像预测相对人的病毒式传播的表现 (60.12%) 要好。使用我们深相对属性, 机器可以更准确地预测相对病毒式传播比人类! 这是因为: (1) 人类不完全了解是什么让一个图像病毒 (因此需要一个研究这样和自动的方法来预测网路效应) 和 (2) 由机器训练有素的属性分类可能已锁存到的病毒含量的偏差。得到的属性了解到概念可以是来自这些属性的人类感知不同。

虽然我们的预测效果很好上述机会, 请注意, 从这些图像中提取的属性是不平凡的, 给定数据集中的图像的多样性。虽然脸检测和动物通常被认为可靠地工作足够的 [16], 记得在 reddit 的图像有挑战性由于其非写实, 嵌入式文字内容和图片组成。为了量化典型的视觉数据集和我们的数据图像的质的区别, 我们训练的分类器将图像分类为属于我们的病毒式传播数据集或数据集 SUN [38, 34]。我们提取从我们的数据集, 并从 SUN 的图像数据相若数目 DECAF6 功能。得到的分类器能够以全新的形象归类为来自两个数据集以 90.38% 的准确度的一个到来, 确认质的差异。此外, 对于普及 [23] 开发的度量应用到我们的数据集中输出偶然类似的结果 (表 2)。因此, 我们的数据集提供了一个新的制度, 以研究图像理解问题。

#### 4.2 附近环境

推理关于图像对如我们上面相对网路效应确实, 导致的图像在其网路效应的人类感知的图像的附近的影响的问题。我们设计了一个实验 AMT 探索这个 (图 8)。回想一下, 在涉及相对网路效应预测前面的实验中, 我们形成了对图像, 其中每对含有病毒和非病毒的形象。现在, 我们添加这些对两个“代理”的图像。这些代理被选择为任一类似于病毒图像, 或对非病毒图像, 或随机地。相似度使用的要点描述 [28] 测量的。4 日和 6 日最相似的图像是从我们的病毒图像数据集 (第 3.2 节) 中选择。我们没有选择最接近的两个图像, 以避免几乎相同的匹配, 并确保任务似乎并不像一个“寻找最奇一出”的任务。我们研究了两种不同的实验环境中这三个条件。首先是那里的工人被要求从他们认为是最病毒最少的病毒全部四个排序图像。在第二个试验设计中, 工人仍然示出的所有四个图像, 但被要求

只注释从原始对的两个图像中的一个比另一个更病毒。也许“代理”的图像仅仅存在影响病毒式传播的看法？

对于这两种情况下，我们仅检查病毒和非病毒性的图像的相对排名。

	Sort 4	Sort 2
Viral-NN	65.16%	66.64%
Non viral-NN	68.60%	65.56%
Random	52.24%	65.00%

表 3: 在不同的代理图像人力排名的准确性。

在六个场景的工人精度表 3 所示。我们可以看到，当记者问到所有四象，排序，确定真正的病毒图像更难随机代理的存在，因为他们往往在预测混淆的工人和他们的表现病毒式传播下降到近的机会。精心挑选的代理的存在，仍然可以使目标图像病毒突出。什么时候

要求只是所感兴趣的两个图像进行排序，性能总体较高（因为任务是不太笨重）。但更重要的是，表现在三个条

件（排序 2）非常相似。这表明，或许代理图像的仅仅存在不影响病毒式传播的预测。

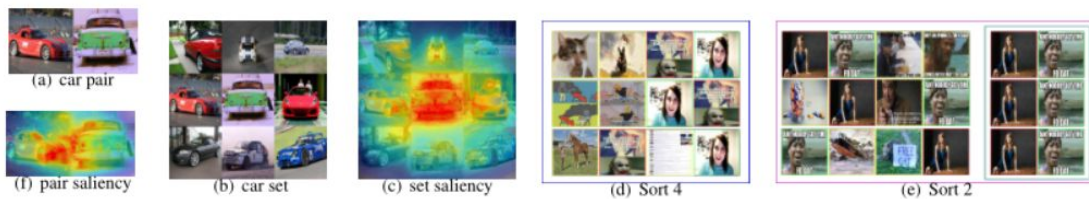


图 8:

的值一车汽车是如何红色，或是否更红比其他如果多个图像被添加到池（a）不改变（b）中。然而，这可能看起来多种病毒的图像 - 通过显着性[22]（例如，红葡萄酒 Ferrari 在（f）的比其他图像时，可能会开始看起来好象比取决于加入到混合物中的图像相同的图像更少的病毒可视化。参照图图 8（c）。在我们的实验中，工人们被要求在一个实验设计（四）在他们的病毒式传播的升序四象排序，而他们被要求只有 2 个在另一个设计（五）排序图像，被显示他们的全部 4 个后。在这两种情况下，也有兴趣只有两个目标图像（病毒：绿色，非病毒：红色），而其他两个图像是代理图像（黄色）加入混合物。这些图像被选择为使得它们在靠近（在要旨空间）对病毒的目标图像（顶行），该非病毒目标图像（中间行），或随机的

（底行）。开发可推理等高阶现象组级别的图像特征并没有得到很好的研究在视觉领域。视觉搜索或显着性进行了研究，以确定哪些图像或图像区域弹出。但在相对排序变革模式同一组基于其他图像的存在图像还没有被开发。这种模型可以让我们选择图像的理想设定包围通过增加其将病毒机会的图像。

#### 4.3 时间背景

在审查中的图像网路效应的空间附近的图像的效果，我们现在研究的时间方面的影响。特别是，我们向用户展示同一对图像的其他后，在 4 个不同分辨率一个在相对网路效应实验中使用在第 4.1.3 节：8×8，16×16，32×32，360×360（原始）。我们选择模糊模拟在缩略图大小的第一印象判断，当图像 are' previewed

“。在每个阶段，我们要求他们该图像他们认为更可能是病毒。病毒式传播的预测表现得 47.08%，49.08%，51.28%和 62.04%。网路效应预测，即使在  $32 \times 32$  的图像，其中，人类已显示识别语义内容图像中非常可靠 [35] 降低到机会。试报告感到惊讶的图像的 65%。我们发现预测的病毒式传播和惊喜之间的相关性 -0.07 真正的病毒式传播和惊喜之间的相关性 -0.04，和。也许人都是坏估计他们是否是真正的惊讶与否，并要求他们未必有效；或者真正的惊喜是不是有病毒式传播相关。

#### 4.4 原文上下文

作为第一个实验来评价图像的标题的作用，我们显示图像的工人对，问他们觉得哪一个更有可能是病毒。然后，我们揭示了图像的标题，并再次问他们同样的问题。我们发现，获得标题勉强提高病毒式传播预测（62.04%对 62.82%）。这表明，也许标题不会动摇科目后，他们已经判断的内容。

我们的第二个实验是相反成立。我们首先表明工作者称号独自一人，问他们哪个称谓更容易使图像是病毒。然后，我们给他们的形象（与冠军一起），问他们同样的问题。工人病毒式传播亲戚预测比单独使用标题（46.68%）的几率更糟糕。有趣的是，已经由标题底漆，即使是具有获得图像的表现没有显著上述改善未经标题（62.04%）被引查看图像时的机会（52.92%），比他们的表现显著降低。这表明，图像内容似乎是在图像网路效应的人类感知的主要信号。但是请注意，这些实验不分析可以嵌入在图像的文字的作用（记因！）。

## 五. 结论

我们研究了从计算机视觉的角度病毒的图像。我们从 reddit 的，世界各地的病毒含量的主要引擎推出了三款新的图像数据集。我们定义使用 reddit

的元数据的病毒式传播的分数。我们发现，网路效应可以更准确地为相对概念预测。虽然人类可以预测从图像内容相对病毒式传播，机器是无法做到使用低级别的功能。高层次的图像理解是关键。我们确定了五个关键的可视属性与病毒式传播相关：动物，合成产生的，（不）美丽的，明确的和性。我们预计采用深图像特征这些相对属性。（：60.12%，比人更高的性能），使用这些深层相对属性的预测为特征，机（SVM）可以的 68.10% 的准确预测病毒式传播。最后，我们研究图片病毒式传播的人预测如何与不同的“上下文”变化 - 内在的，空间（附近），时间和文本。这项工作是在了解图像内容操作复杂，但重要的现象的第一步。我们已经表明，需要先进的图像理解来预测病毒式传播，以及我们的数据集和典型的视觉数据集之间的质的区别。这开辟了视觉领域的新机遇。我们的数据集和注释将被公布于众。

## 六. 致谢

这项工作是部分由 ARO YIP65359NSYIP 支持和 NSF IIS-1115719。我们还要感谢斯坦尼斯 Antol，迈克尔科格斯韦尔，恶劣的阿格拉瓦尔和阿琼 Chandrasekaran 他们的反馈和支持。

### References

- [1] H. Agrawal, N. Chavali, M. C., Y. Goyal, A. Alfadda, , P. Banik., and D. Batra. Cloudcv: Large-scale distributed computer vision as a cloud service, 2013. 4
- [2] A.-L. Barabasi. The origin of bursts and heavy tails in human dynamics. Nature, 2005. 1
- [3] A. Berg, T. Berg, H. Daume, J. Dodge, A. Goyal, X. Han, A. Mensch, M. Mitchell, A. Sood, K. Stratos, et al. Understanding and predicting

- importance in images. In CVPR, 2012. 2
- [4] J. Berger. Arousal increases social transmission of information. *Psychological science*, 2011. 1, 2, 6
- [5] J. Berger. *Contagious: Why Things Catch On*. Simon & Schuster, 2013. 1, 2
- [6] J. Berger and K. L. Milkman. What makes online content viral? *Journal of Marketing Research*, 2012. 2, 3, 4
- [7] J. Berger and E. M. Schwartz. What drives immediate and ongoing word of mouth? *Journal of Marketing Research*, 2011. 1
- [8] A. Beutel, B. A. Prakash, R. Rosenfeld, and C. Faloutsos. Interacting viruses in networks: can both survive? In SIGKDD, 2012. 2
- [9] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar. Face swapping: automatically replacing faces in photographs. In *Transactions on Graphics (TOG)*, 2008. 4
- [10] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 223 - 232. ACM, 2013. 2
- [11] Z. Chen and J. Berger. When, why, and how controversy causes conversation. *The Wharton School Research Paper*, 2012. 1
- [12] E. J. Crowley and A. Zisserman. In search of art. In *Workshop on Computer Vision for Art Analysis, ECCV*, 2014. 2
- [13] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005. 4
- [14] S. Dhar, V. Ordonez, and T. L. Berg. High level describable attributes for predicting aesthetics and interestingness. In CVPR, 2011. 2
- [15] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. *arXiv preprint arXiv:1310.1531*, 2013. 4, 7
- [16] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *arXiv preprint arXiv:1311.2524*, 2013. 7
- [17] M. Guerini and J. Staiano. Deep feelings: A massive cross-lingual study on the relation between emotions and virality. *arXiv preprint arXiv:1503.04723*, 2015. 2, 6
- [18] M. Guerini, J. Staiano, and D. Albanese. Exploring image virality in google plus. In *Social Computing (SocialCom), 2013 International Conference on*, pages 671 - 678. IEEE, 2013. 2, 3
- [19] P. Isola, D. Parikh, A. Torralba, and A. Oliva. Understanding the intrinsic memorability of images. In NIPS, 2011. 2, 6

- [20] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In CVPR, 2011. 2
- [21] E. Johns, O. Mac Aodha, and G. J. Brostow. Becoming the Expert – Interactive Multi-Class Machine Teaching. In CVPR, 2015. 2
- [22] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In Computer Vision, 2009 IEEE 12th international conference on, pages 2106 – 2113. IEEE, 2009. 8
- [23] A. Khosla, A. D. Sarma, and R. Hamid. What makes an image popular? In International World Wide Web Conference (WWW), Seoul, Korea, April 2014. 2, 3, 7
- [24] C. Kiddon and Y. Brun. That’s what she said: Double entendres identification. In ACL (Short Papers), 2011. 4
- [25] H. Lakkaraju, J. McAuley, and J. Leskovec. What’s in a name? understanding the interplay between titles, content, and communities in social media. ICWSM, 2013. 1, 2, 3
- [26] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. Transactions on the Web, 2007. 1, 2, 4
- [27] M. Nagarajan, H. Purohit, and A. P. Sheth. A qualitative examination of topical tweet and retweet practices. In ICWSM, 2010. 2
- [28] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. IJCV, 2001. 4, 7
- [29] D. Parikh and K. Grauman. Relative attributes. In ICCV, 2011. 5
- [30] P. Shakarian, S. Eyre, and D. Paulo. A scalable heuristic for viral marketing under the tipping model, 2013. 1, 2
- [31] M. Spain and P. Perona. Measuring and predicting object importance. IJCV, 2011. 2
- [32] B. Suh, L. Hong, P. Pirolli, and E. H. Chi. Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In Social Computing, 2010. 2
- [33] A. Torralba. Contextual priming for object detection. IJCV, 2003. 4
- [34] A. Torralba and A. A. Efros. Unbiased look at dataset bias. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 1521 – 1528. IEEE, 2011. 7
- [35] A. Torralba, R. Fergus, and W. T. Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. TPAMI, 2008. 4, 8
- [36] N. Turakhia and D. Parikh. Attribute dominance: What pops out? In ICCV, 2013. 2
- [37] W. Y. Wang and M. Wen. I can has cheezburger? a nonparametric approach to combining textual and visual information for predicting and generating popular meme



descriptions. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015. 2

[38] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In CVPR, 2010. 4, 7

