**学习视觉注意以识别自闭症谱系障碍患者**

**摘要**

本文提出了一种利用眼动追踪和深度神经网络定量客观诊断自闭症谱系障碍（ASD）的新方法。泛自闭症障碍很普遍，在美国有 1.5% 的人。缺乏用于早期诊断的临床资源一直是一个长期存在的问题。这项工作具有三个独特的特征：首先，所提出的方法是数据驱动的，没有假设，这对于理解ASD和其他神经发育障碍的新发现非常重要。其次，我们将分析集中在健康人和ASD患者之间眼球运动模式的差异上。基于Fisher评分的图像选择方法允许使用最具鉴别性的内容进行特征学习，从而实现高效准确的诊断。第三，我们利用深度神经网络的最新进展进行预测和可视化。实验结果表明，在诊断测试中使用的多个评估指标方面，该方法具有优越的性能。

1. 介绍

神经发育障碍的“金标准”诊断过程是昂贵、主观且耗时的。依靠神经科医生，儿科医生，病理学家和治疗师的多学科团队，通常需要多年的连续访谈和行为评估，这些都容易受到解释性偏见的影响。例如，自闭症谱系障碍（ASD）影响美国每68人中就有一人[9]。由于ASD的患病率和有限的临床资源，早期诊断和干预措施无法广泛获得或适用。光谱上的高功能人群在童年时期通常被忽视。因此，几十年来，定量和客观的诊断工具一直非常需要，但进展甚微。

最近，我们目睹了两个相关且重要的领域的快速增长，这两个领域可能会彻底改变神经发育障碍的诊断 - 视觉注意和人工神经网络（DNN）。注意力网络在大脑中如此普遍，以至于许多神经发育障碍与对视觉刺激的非典型注意力有关[4]。然而，目前注意力的神经行为研究依赖于人为的视觉刺激和针对特定假设的结构化实验室任务，这限制了这些研究的普遍性。

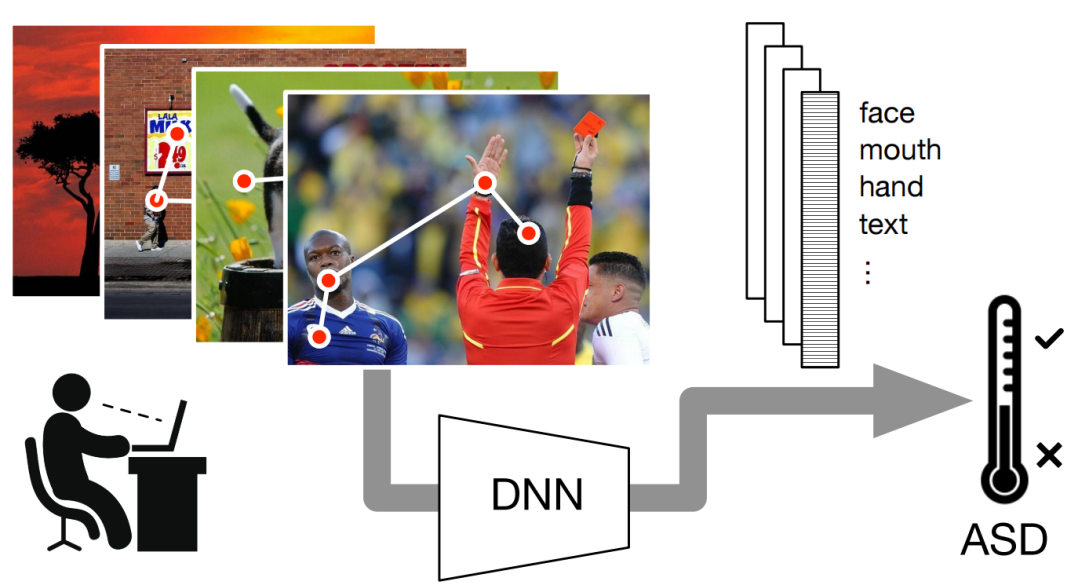


图 1概述我们的实验和分类范式。受试者自由观察一系列自然场景图像，同时记录他们的眼球运动。编码社会和非社会信息的判别特征是从视觉注意力的深度神经网络中提取的，并跨图像集成以预测受试者的ASD风险。

为了解决ASD诊断中长期存在的问题，我们建议利用视觉注意力在大脑中的普遍性和DNN的学习能力。在DNN的推动下，注意力的计算模型近年来取得了显着进展，编码了丰富的语义，并在复杂的自然场景中很好地预测了人类的注意力。由于ASD对社会刺激表现出非典型的凝视模式[32，13，31]，DNN编码高级语义和社会内容的能力特别适合于识别ASD患者。

具体来说，我们提出了一种具有自然场景观看的新诊断范式。使用DNN模型记录受试者的眼球运动并进行定量分析。为了有效和高效的诊断，我们的方法通过对Fisher分数进行排名来选择最具鉴别性的图像，并通过创建和分析眼睛注视的差异图来学习最具鉴别性的特征。通过使用自然场景刺激和 DNN，诊断过程完全由数据驱动且无假设。

综上所述，本文有三大贡献：

首先，由于自由观看任务，刺激和数据驱动模型的普遍性，我们的方法很容易适用于光谱上的大多数人，并且可以推广到其他神经发育障碍。

其次，选择判别图像和学习判别特征可提高ASD诊断的准确性，同时减少诊断所需的数据量和时间。

最后，我们基于DNN的模型在临床人群分类方面表现出卓越的性能。DNN可视化有助于丰富我们对ASD中非典型注意力的理解。

1. 相关工作

机器学习方法已应用于各种形式的自闭症研究，例如基于商的ASD诊断工具[2]，磁共振成像数据[11]，早期语言的声学数据[25]和运动学数据[5]。然而，这些研究中的大多数都依赖于手工制作的特征和简化的线性模型，这些模型通常无法在控制较少的环境中处理人类行为的复杂性。在这里，我们简要回顾了使用机器学习分析眼动追踪数据以进行自闭症研究和诊断的最相关的工作。

眼球运动编码有关个体的注意力和动眼运动控制的丰富信息，这可能有助于表征ASD特征。与标准诊断商相比，眼球运动的自动和定量处理将潜在地导致更客观和可及的ASD诊断。近日，Liu等人[22]分析了ASD患儿在人脸识别任务中的凝视模式。他们提出了一种机器学习方法，对29名患有ASD的儿童和两组匹配的对照组进行分类。尽管它们具有突出的准确性，但面部刺激和结构化识别任务高度依赖于有关ASD的现有知识，从而限制了它们对其他临床人群或幼儿的推广，他们可能不理解或遵守任务指令。

为了对患有各种神经发育障碍（即注意力缺陷多动障碍，胎儿酒精谱系障碍和帕金森病）的临床人群进行分类，请注意，这项工作没有研究ASD），Tseng等人[35]分析了观看短视频剪辑时的凝视模式。他们将凝视模式与视觉注意力计算模型的特征相结合，并展示了在识别特定疾病时结合注意力相关特征的优势。他们的注意力模型只关注早期的视觉显著性，但它没有考虑到可能更显着地影响视觉注意力的高级语义和社会信息。

我们的工作主要受到Wang等人的综合研究的启发[38]。他们使用线性支持向量机（SVM）和三个级别的特征量化了患有ASD的高功能成年人的视觉注意力。通过记录的眼睛固定来学习特征权重，并用于量化ASD患者和健康对照组之间的不同注意力偏好。这种方法的主要局限性是手工制作的特征和对感兴趣对象进行手动注释的要求。此外，Wang等人[38]只研究了群体之间的差异，而没有对个体受试者进行分类。相反，为了解决这些问题，我们的方法使用深度神经网络（DNN）自动从自然场景中提取图像特征，并使用这些特征将ASD患者与健康对照区分开来。

深度神经网络是机器学习中最先进的架构，已广泛应用于各种计算机视觉任务。有了足够的训练数据，DNN 模型在视觉识别、检测、分割等方面的表现就压倒性地优于手工制作的功能。深度视觉注意网络也证明了预测人类外观的有效性[37，18，21，14，27，20]。一些最先进的模型将从对象识别任务中学到的知识转移到注意力预测问题[18]或使用显着性评估指标作为目标[14，20]。与专注于预测一个种群的凝视分布的相关工作不同，本研究首次开发了一种DNN模型来预测两个种群之间的差异。用突出显示组间差异的概率图替换传统的注视基实，使 DNN 能够学习对一组人显著但对另一组人来说不显著的判别特征。

1. 方法

在这项工作中，我们利用从一系列判别图像中学到的丰富DNN特征来对个体的组成员身份进行分类。图 2 演示了我们基于 DNN 的方法的概述。我们方法的关键组成部分包括眼动追踪数据收集，判别图像的选择，基于DNN的特征学习和基于SVM的分类。

3.1.眼动追踪

Wang等人收集了眼动追踪数据[38]，来自20名患有ASD的高功能成年人和19名神经和精神健康对照组，用于自由观看任务。受试者的视力正常或矫正至正常。两组受试者在性别，年龄，种族和教育程度上进行了匹配。当线人可用时，ASD组均符合自闭症诊断观察时间表（ADOS）[23]和自闭症诊断访谈修订版（ADI-R）[24]或社会沟通问卷（SCQ）的截止分数。

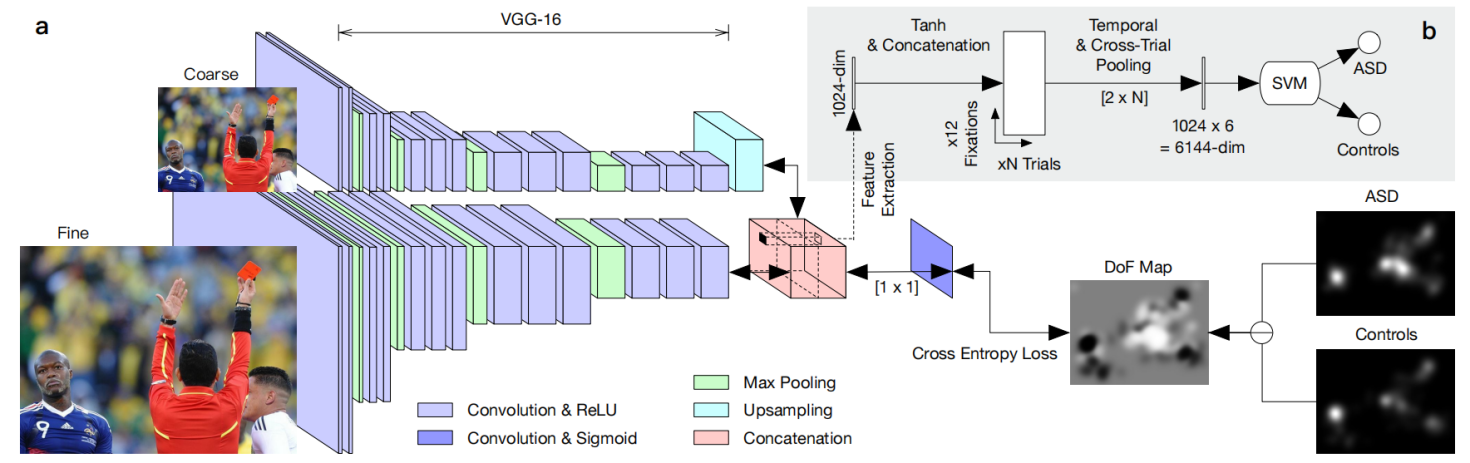


图 2概述了所提出的特征学习和分类方法。（a）识别性的图像特征被端到端地学习以预测注视地图的差异。（b）在N次试验中提取并整合固定像素处的特征，用SVM对每只眼睛的行为进行分类。

受试者观察了来自OSIE [39]数据库的700张图像。由于自然场景中的对象种类繁多，这些图像在理解视觉注意力的不同方面（从低级特征对比度到高级语义）方面显示出优势。在眼动追踪实验中，图像被洗牌并在连续的实验试验中呈现。每次试验都以漂移校正开始，受试者将一个点固定在空白屏幕的中心，然后按下按钮继续。漂移校正后，目标图像呈现三秒钟。受试者被指示自由观看图像。在图像查看期间，使用非侵入性红外遥控器Tobii X300眼动仪以300 Hz记录双眼眼球运动。该实验分为七个记录会话，每个会话100个试验，在会话之间有短暂的休息和重新校准。受试者在两小时内完成了实验。

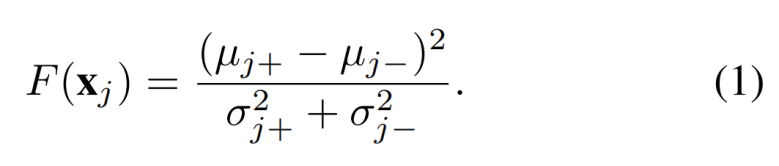
我们处理了原始眼动追踪数据以进行进一步分析。使用Cluster Fix [17] 从原始数据中检测到注视点和扫视 - 一种非参数注视检测算法。我们将每个扫视-注视对视为视觉扫描过程的时间注视阶段。对于所有试验，我们在第十二个固定阶段后丢弃了数据。

请注意，眼动追踪精度可能会受到校准误差、环境条件、双目视差和受试者表现的影响。例如，在[38]中，ASD的平均眼动追踪误差约为40像素，控件的平均眼动追踪误差约为18像素。因此，Wang等人[38]丢弃了不太准确的眼睛的数据。然而，在这项工作中，我们分析了双眼的数据，防止了模型过度拟合到显眼，并提高了其对眼动追踪误差的鲁棒性。此外，在我们的实验中，我们将受试者的两只眼睛分类为两个单独的数据样本。因此，每只眼睛都是独立分类的，即使只有单目数据可用，这也将允许未来的受试者进行测试。

* 1. 图像选择

OSIE[39]数据库最初是为视觉注意力的一般研究而设计的，但不是针对自闭症的。由于实验试验的数量可以直接确定诊断测试的成本，因此我们从原始数据库中选择了一个图像子集，可以最好地区分ASD患者和对照组之间的凝视模式。我们根据Fisher评分方法[10]进行了特征选择，以找到最具鉴别性的图像。

从700张自然景图像中选择一个判别子集是我们方法的重要组成部分。判别性图像应最大限度地揭示来自不同组的受试者之间的统计差异，同时对组内变异的敏感性最小。Fisher评分的基本思想是找到对受试者组分类最重要的特征子集。选择这些特征的原理是，位于同一类中的数据点应彼此靠近，而位于不同类中的数据点应彼此相距较远。具体而言，考虑一个特征向量 x，它表示个体的注视模式，其 j 次维数表示为 xj 。设 （μj+， σj2+） 和 （μj−， σj2−） 分别表示 xj 在 ASD 和对照组之间的均值和方差。xj 的 Fisher 分数计算如下：

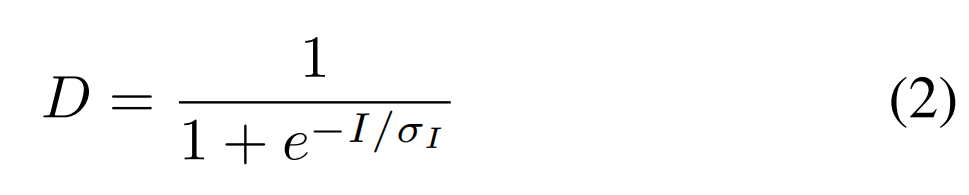


为了执行图像选择，定义了一组凝视特征，描述了受试者的观看行为。在每个注视阶段，我们计算了注视点的三个空间特征（即它们的水平和垂直位置，以及与屏幕中心的距离）和四个动眼特征（即注视持续时间，扫视幅度，扫视持续时间和扫视速度）。这些特征向量在所有注视阶段和图像上串联，代表了个体在眼动追踪实验中的观察行为。然后在所有特征尺寸上计算Fisher分数，并按图像求平均值。我们选择了得分最高的图像子集，并仅对所选图像进行特征学习和分类。

* 1. 从注视图的差异中学习判别特征

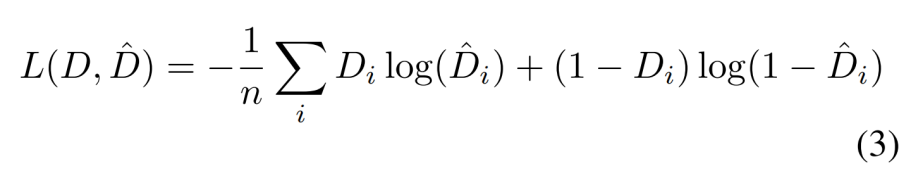
用于预测人类视觉注意力的基于学习的模型通常经过优化，以区分场景中的突出和非关注元素。它们的基本事实是眼睛注视的概率分布，即注视图。由于我们的目标是通过两个临床人群的注视点来区分它们，因此我们的网络根据注视图（DoF）图的差异进行了优化，突出了两个注视图之间的细微差异。如图2a所示，DoF图中的值越高，表示ASD组的注视点越多，而较低的值表示对照组的注视点越多。

具体而言，通过将固定像素设置为1，将其他像素设置为零，计算每个总体的所有注视点的积分来计算注视图。这两个映射使用高斯核（带宽 = 1◦ ）进行平滑处理，并通过总和进行归一化。设 I+ 和 I− 分别表示 ASD 组和对照组的最终注视图。图像的 DoF 映射可以计算为



其中 I = I+ − I− 是两个映射的像素减法，σI 表示 I 的标准差。

如图2a所示，我们的网络架构遵循SALICON网络[14]的设计，这是图像显著性预测领域的最先进技术之一。它由两个并行的VGG-16网络[33]组成，以两种不同的空间分辨率处理输入图像。精细比例网络将原始图像作为输入，而粗略比例输入是图像下采样一半。从VGG-16中移除完全连接的层后，两个网络的最后一个卷积层的激活将缩放到相同的大小并连接。在我们的实现中，添加了一个新的卷积层，然后是Sigmoid激活以预测DoF映射。卷积层的内核大小为 1×1，不带填充，表示加权线性组合。模型参数在端到端反向传播中进行了优化，并具有像素级交叉熵损失:



其中，求和在所有 n 个输出像素上。D 和 ˆD 分别是预测的 DoF 映射和地面实况。Di 和 ˆDi 是它们在第 i 个像素处的值。VGG-16 层的参数初始化为ImageNet 上的预训练参数 [8]。然后，在当前数据库上对网络进行微调。

* 1. 分类

我们在每个注视位置提取了习得的DNN特征，并将其整合到所有实验试验中。利用这些特征，我们训练了一个线性SVM来对两个群体进行分类。

具体而言，如图2b所示，我们从1024-通道串联激活图中提取了固定位置处的DNN响应，从而在每个固定点处产生了1024维特征向量。然后，对提取的特征应用tanh函数以转换[0，1]内的响应。对于每个眼球运动扫描路径，我们按时间顺序连接所有注视点的特征向量，从第一个注视点到最后一个注视点，如果注视点少于十二个，则将零附加到末尾。这允许对注意力随时间的动态变化进行建模。为了减少维度和控制过拟合，我们对每两个连续的注视点（即，size=2，stride=2）进行局部平均池化。最后，对所有试验的特征进行平均，以代表每只眼睛的整体注意力模式。

SVM使用提取的特征进行训练，以找到线性决策边界，其中两个群体之间具有最大余量。在测试阶段，学习的SVM模型对每只眼睛注视数据的组成员身份进行了分类，并提供了相应的置信度评分，这使我们能够引入灵活的阈值来确定最终的分类标签。

1. 实验结果

在本节中，我们将报告实施细节并对所提出的方法进行全面评估。

4.1. 实现

培训和测试。分类的绩效是通过一个受试者的遗漏交叉验证来评估的。留出一个交叉验证已被广泛用于机器学习，因为它能够返回误差概率的几乎无偏估计[36]。特别是在这项研究中，由于每个受试者的双眼眼球运动将被单独分类，为了测试我们数据库中来自39名受试者的78个数据点，因此进行39轮分类就足够了，每个测试有两个数据点，并与剩余的76个数据点一起训练。图像选择、DNN 和 SVM 参数均由交叉验证决定。

在每次交叉验证中，根据其Fisher分数的排名选择100张图像的子集。DNN 和 SVM 分类器在选定的 100 张图像上进行训练和测试。如此小的一组刺激可以显着降低临床应用中眼动追踪实验的成本，因为只有选定的图像才能用于测试。

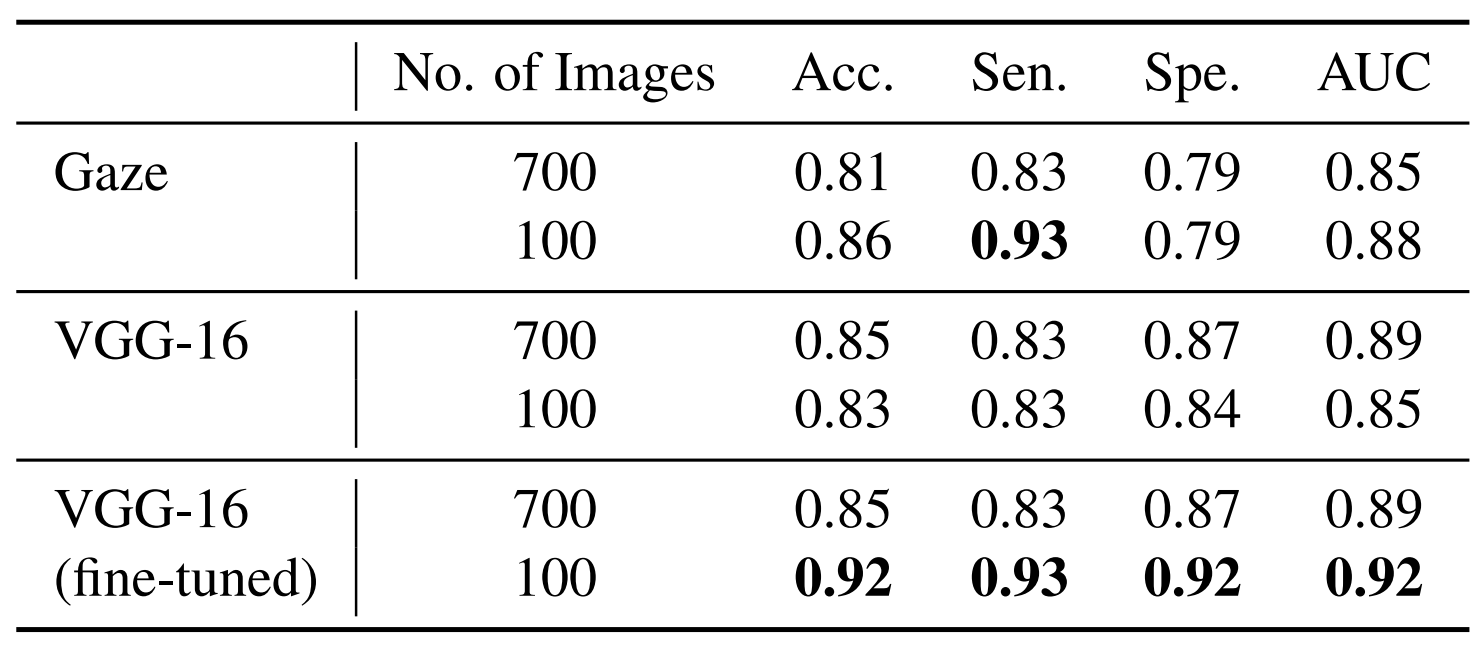
我们的模型是用Python语言实现的，带有Caffe和Scikit学习库。在神经网络训练期间，我们使用随机梯度下降和一幅图像的批量大小对网络进行了微调，总共进行了10000次迭代。基本学习率为10%−当误差稳定时，6除以10。我们使用了10的重量衰减−5和0.9的动量。按照【14】中的实践，未使用辍学，也未进行数据扩充。为了监控收敛和过度拟合，每1000次迭代后对网络进行交叉验证。在GTX Titan GPU上训练网络大约需要两个小时。对于SVM分类，我们使用惩罚参数C=1的L2正则化。

评估指标。我们从准确性，特异性和灵敏度方面定量测量了我们提出的方法的性能。这些指标已被广泛用于几乎所有类型的临床评估。准确性测量ASD组和对照组中正确分类样本的速率。敏感性（即真阳性率）和特异性（即真阴性率）分别测量ASD组和对照组中正确分类的样本的速率。

我们还通过引入灵活的分类阈值进行了接收器工作特性（ROC）分析。SVM置信度评分高于阈值的测试样本被给予阳性ASD诊断。通过改变阈值，我们绘制了所有真阳性率与假阳性率的ROC曲线，并计算了ROC曲线（AUC）下的面积作为分类性能的定量度量。

* 1. 分类性能

表 1使用不同的注视和 DNN 特征集，比较完整数据库和所选 100 张图像上的分类性能。



本文在表1中给出了所提方法的定量结果。为了研究所提出的DNN特征的影响，我们使用三组不同的特征来训练和评估SVM分类器，包括（1）图像选择中使用的凝视特征（参见第3.2节），（2）微调前的VGG-16特征，以及（3）微调后的VGG-16特征。为了演示基于Fisher评分的图像选择的效果，我们还比较了图像选择前后的表现。

如表1所示，对所选图像上的网络进行微调可显著提高分类精度。此外，图像选择将凝视特征的灵敏度提高了10%，这表明ASD组的数据方差减小，使其与对照组更加明显。此外，微调将特异性提高了8%，这表明DNN编码了对照组一致固定的视觉刺激。换句话说，这些控件可以通过他们所看到的东西来更好地区分，而不是他们如何转移视线。有趣的是，微调只对选定的图像有效。很可能只有在Fisher得分较高的图像中，两组受试者才关注不同的图像特征，这些特征最有效地将ASD患者与健康对照组分开。

由于我们是第一个在ASD诊断的自由观察任务中使用眼动追踪的人，因此没有相关的工作可以直接比较。虽然针对不同的受试者群体，但Liu等人[22]在方式，实验环境和受试者数量方面可能是与我们的工作最相似的。总体而言，如表2所示，我们的方法达到了显着的分类性能，优于[22]和许多基于商的诊断工具[12]，包括ADOS，儿童自闭症评定量表（CARS）和自闭症谱系障碍 - 儿童诊断（ASDDC）。值得注意的是，我们的方法只需要大约10分钟的眼动追踪，与[22]相当，需要执行人脸识别任务，而其他方法则需要10到45分钟。通过免费图像查看任务，我们的方法证明了准确识别ASD患者的可行性和可推广性。

我们的表现与传统的基于商的自闭症自我评估方法的结果一致。我们将SVM预测的置信度评分与受试者的临床特征进行了比较。对同一主题的双目预测进行了平均。由于缺乏对照的ADOS和ADI-R评估，我们仅将预测与成人ASD特征的两种定量测量结果进行比较，即自闭症谱系商数（AQ）[1]和社会反应能力量表-2成人形式自我报告（SRS）。

表 2与最相关的眼部追踪(ET)研究和目前的诊断工具进行定量比较。

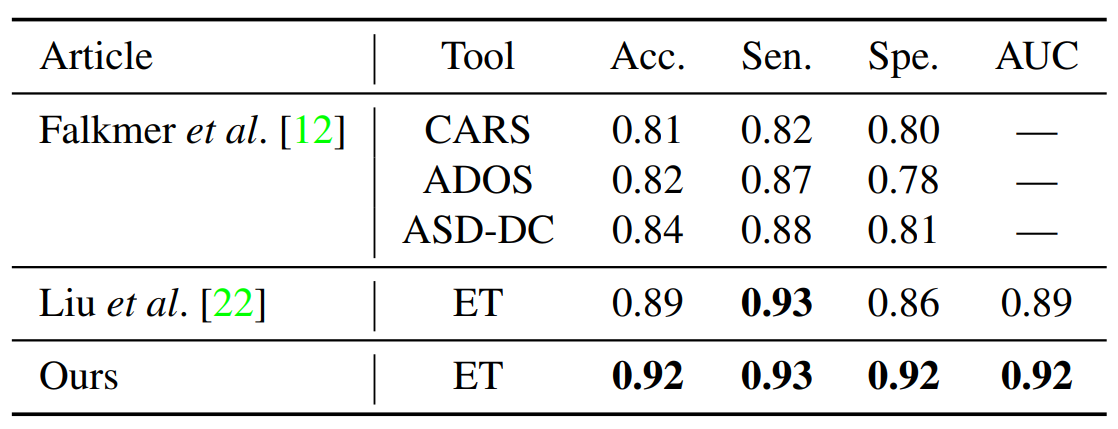
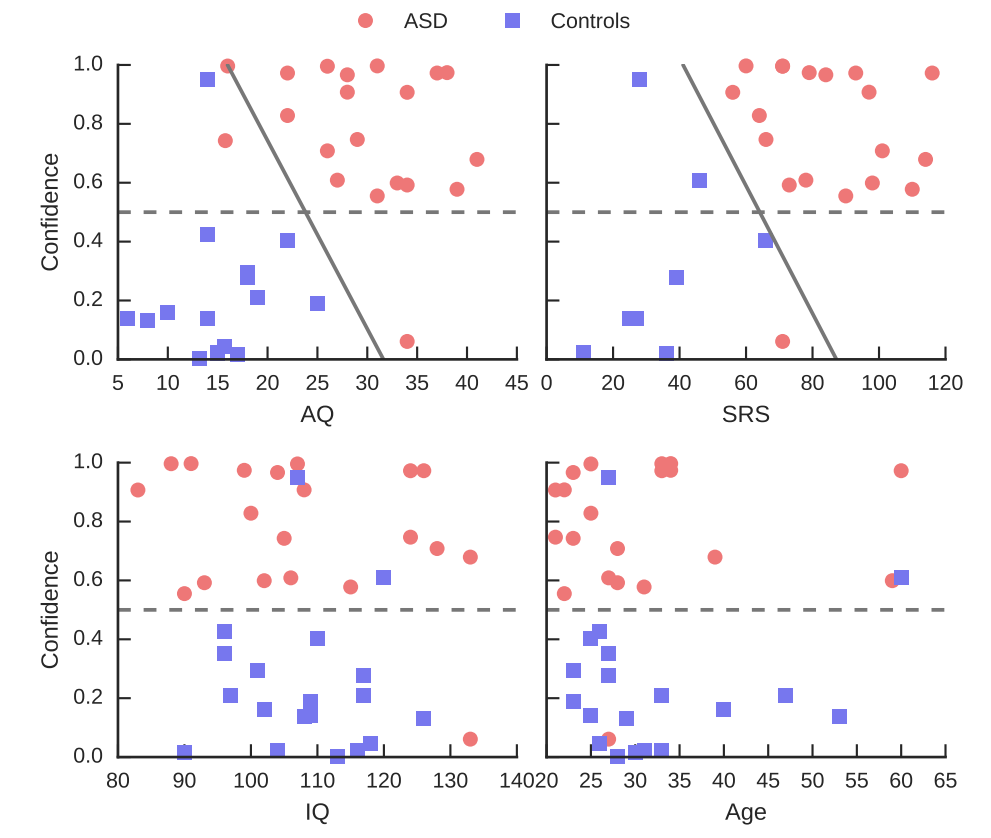


表 3分类置信度与受试者的AQ和SRS分数密切相关，但与智商或年龄无关。虚线表示0.5处的分类边界。实线表示结合模型预测和基于商的评价的最优分类边界。



（有道翻译）

两者都是被广泛认可的成人ASD自我评估工具。我们观察到,分类信心很强的相关性与科目的分数的AQ(皮尔森的ρ= 0.53,p = 0.0084)和SRS(皮尔森的ρ= 0.51,p = 0.0063),但没有相关性与受试者的年龄(皮尔森的ρ= - 0.0035,p = 0.98)或其全面智商分数(智商;ρ = -0.15, p = 0.37)。我们的方法还显示出了补充商的潜力，提高了分类的准确性(见图3)。

* 1. 图像选择

图像选择对分类性能有效（见表1）。下面，我们将通过详细分析进一步阐述其影响。

我们首先研究了七种凝视特征在图像选择中的作用。图4显示了不同注视阶段的Fisher评分变化。在这里，特征分数是图像间的平均值，但不是每个图像内的平均值。我们可以观察到，与四个动眼神经特征（即注视持续时间、眼跳幅度、眼跳持续时间和眼跳速度）相比，在图像选择后，三个空间特征（即水平和垂直位置、到中心的距离）的Fisher得分增加更多。特别是，到中心的距离编码了所有注视阶段中差异最大的信息。这一观察结果证实了之前的发现，即自闭症患者在眼睛注视方面有更强的中心偏好[38]。此外，尽管动眼神经特征的重要性相对较低，但它们在早期注视阶段发挥了重要作用，这可能表明注意力的分离受损，从而阻止了自闭症患者从最初的注视中扫视[7]。

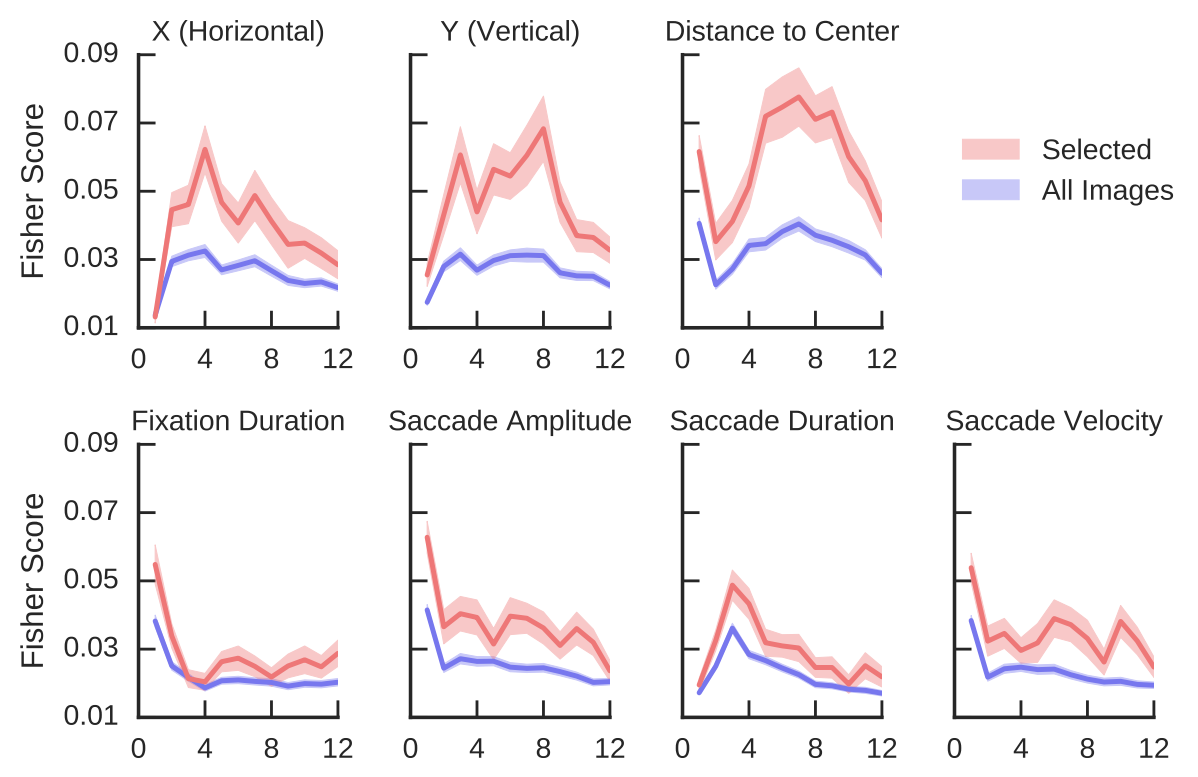


图 3费舍尔对不同时间的凝视特征进行了评分，并将其平均化。水平轴表示固定顺序。误差带表示平均值的标准误差。

我们将Fisher评分与经典显著性评估指标进行比较：AUC[34]、随机AUC（SAUC）[40]、线性相关系数（CC）[26]、信息增益（IG）[19,20]、Kullback-Leibler散度（KL）[15]、标准化扫描路径显著性（NSS）[29]和相似性（SIM）[16]。这些指标通常用于比较眼睛固定的位置或分布[3]。一般来说，Fisher评分与所有具有显著相关性的比较显著性指标一致。然而，与传统的显著性度量相比，它有一些关键的优势。Fisher得分不仅测量组间差异，还对组内差异进行折扣。此外，在这项工作中，明确包括动眼神经特征，这是其他指标无法衡量的。因此，提出的图像选择方法能够在更小的图像集上保持较高的分类性能，而其他指标则失败（见图5）。

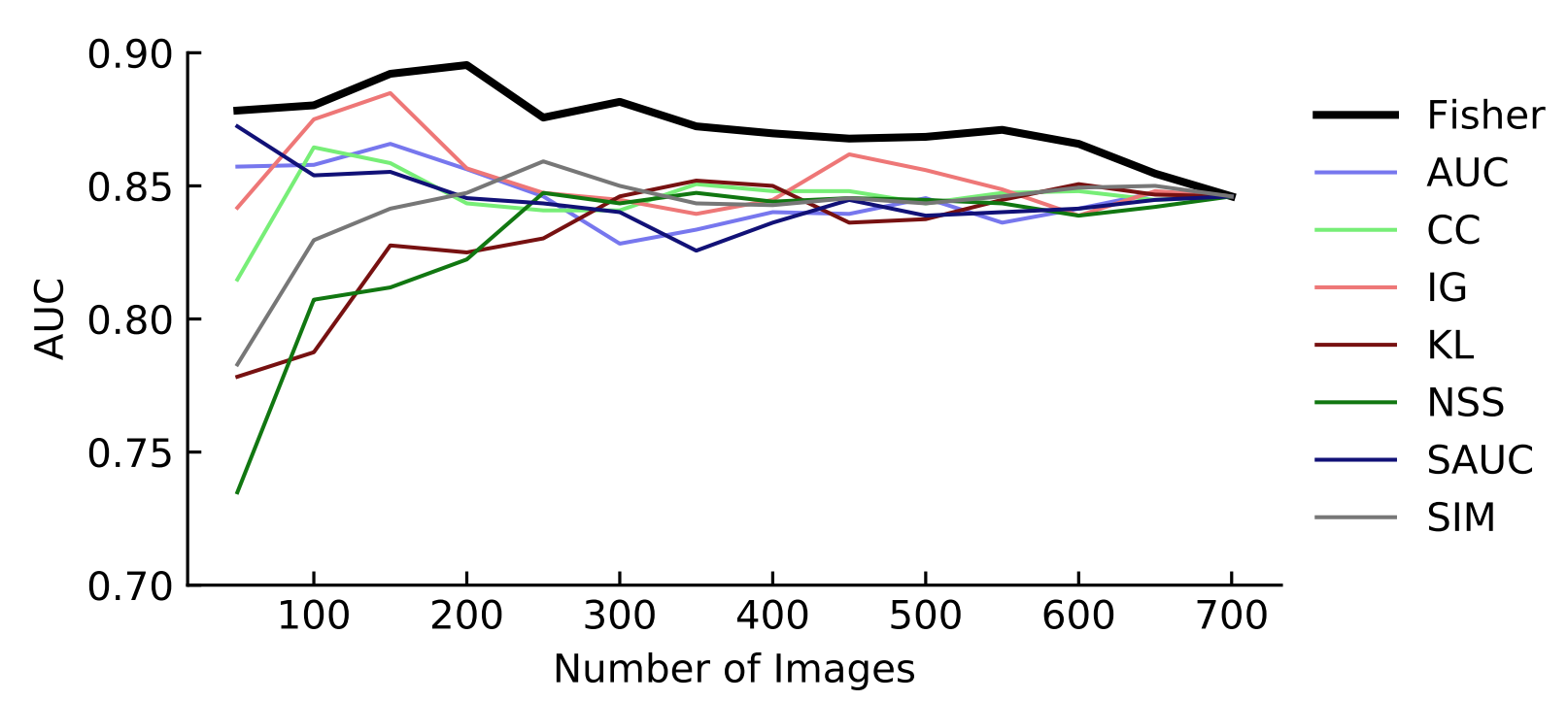


图 4比较不同图像选择方法的分类性能（AUC）。Fisher分数是使用第3.2节中介绍的凝视特征计算的，而其他分数是基于位置或分布的显著性度量。

图6展示了Fisher分数高低图像的定性比较。Fisher得分较高的图像（图6a和6b）在两个群体之间的固定图上有明显差异。特别是，自闭症组始终专注于非社会刺激（即食物和椅子），而对照组则专注于社会刺激（即面部和文本）。对于Fisher分数较低的图像（图6c和6d），两张固定图非常相似。用一个减去另一个只会增加眼睛跟踪误差和背景中的随机固定。我们还观察到，Fisher评分较高的图像往往会吸引更多的定影（Pearson'sρ=0.11，p=0.005），并且定影较少偏向图像中心（Pearson'sρ=0.15，p<0.001）。正如Wang等人[38]所建议的那样，自闭症患者倾向于更少的注视和更强的中心偏好，我们相信这种差异也反映在Fisher评分中。定性和定量评估都表明，所提出的图像选择在我们的方法中起着重要作用，它成功地保留了最具ASD辨别力的特征。

总之，我们的研究结果表明，一小部分自然场景图像可以可靠地识别出患有复杂和异质性神经发育障碍的个体。因此，诊断会话可以在可行的时间长度内进行，并且计算模型可以很容易地部署在临床或家庭环境中。

* 1. DNN特征的可视化

对DoF图上的视觉注意深度神经网络进行微调，可以产生高度区分的特征，显著提高SVM分类性能。在接下来的分析中，我们将学习到的特征可视化，并在现有的自闭症研究中寻找对这些特征及其作用的解释。

我们通过比较学习特征在线性SVM分类中的权重来研究其重要性。图7显示了权重最高和最低的特征（在固定过程中的平均值）。具有正权重和负权重的特征分别支持ASD患者和对照组的分类。为了可视化DNN特征，对于每个通道，我们在六个青少年注视点提供了128×128个图像块，这些图像块在整个过程中产生了最高的响应

相应的神经元。

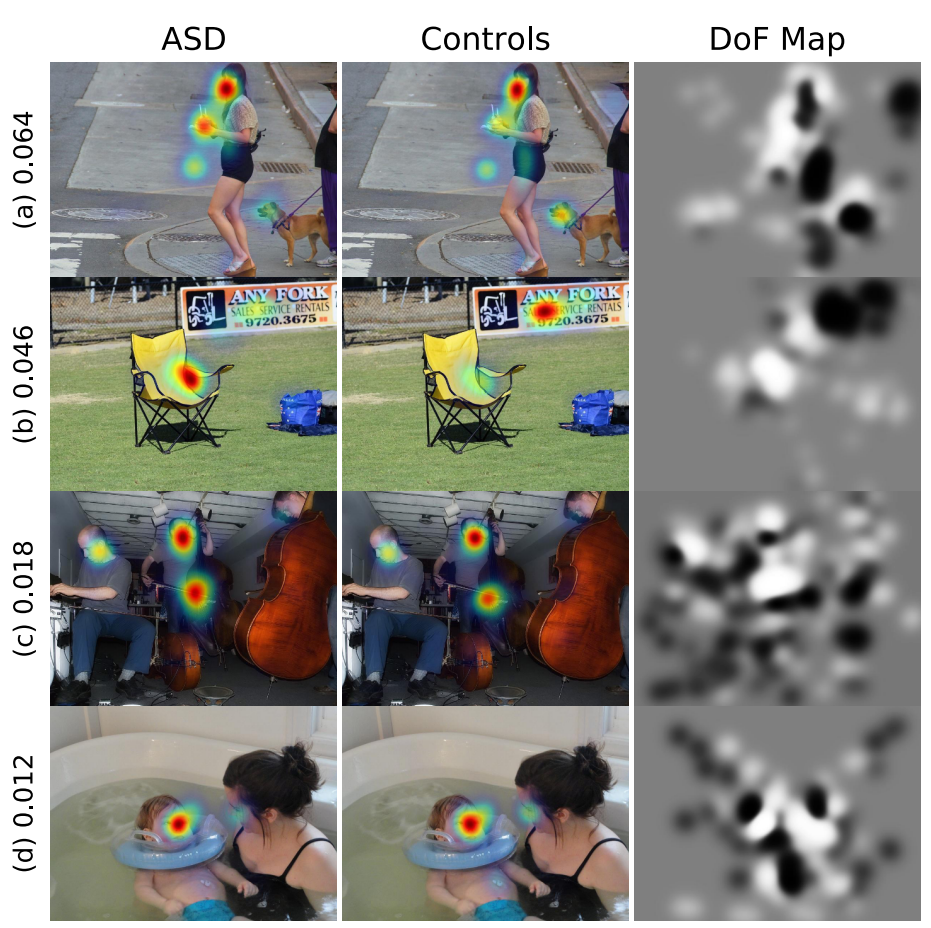


图 5带有Fisher分数的示例图像。ASD组和对照组的固定图叠加在图像上。它们的自由度图以灰度显示。

如图7所示，我们观察到ASD患者的低水平显著性增加，但社会关注度降低。自闭症患者的主要特征是各种非社交对象的中层结构，例如边缘、形状和重复模式（图7a-7d）。虽然他们也会观察面部，但他们会将注意力转向头发和下巴（图7e和7f），这表明他们倾向于避免目光接触[28,6,30]。相比之下，文本、符号、人脸、上身、肢体等社交特征（图7g-7l）识别了对照组。我们还研究了学习到的特征如何与[39]定义的语义属性相对应。值得注意的是，图7g、7i和7l代表的是被OISE数据库中定义的三种社会属性（即文本、面部和凝视）所激活的神经元[39]。这些观察结果虽然是从自然场景图像中获得的，没有任何预先假设，但支持ASD的非社会偏好[32,31]。

与大多数计算机视觉应用程序不同，在这些应用程序中，训练和测试是在不同的图像上进行的，我们的问题是测试新的主题。因此，相同的图像被用于训练和测试，但具有不同受试者的眼球跟踪数据。尽管如此，我们还是进行了一个有趣的实验，来检查网络如何预测新图像的自由度图。这样一个网络将对各种现实世界的应用产生深远的影响，比如设计类似自闭症的书籍或海报。如图8所示，与实际情况相符。与持续关注社会刺激的对照组相比，自闭症组的注意力分配更随机，更难预测。

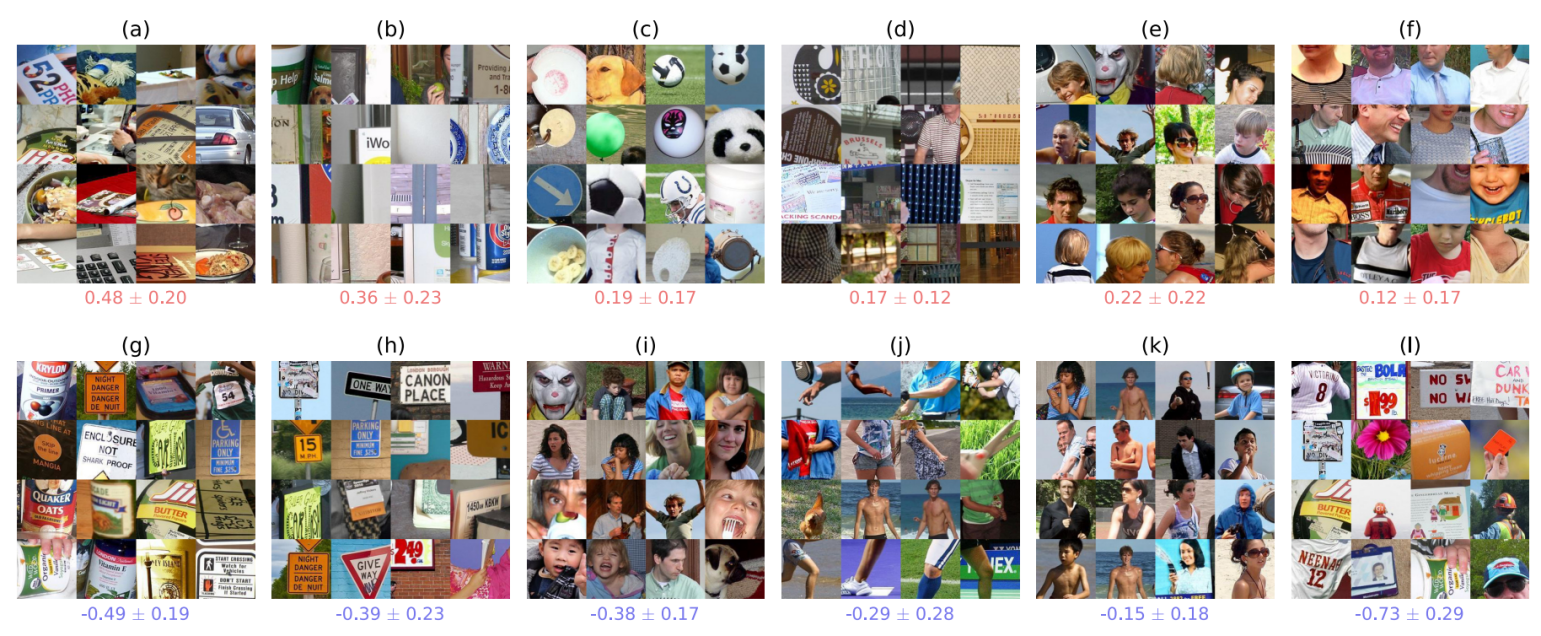


图 6具有最高和最低SVM权重（×10）的特征−3，平均值±标准差），可视化为图像块阵列。正权重和负权重分别代表ASD组和对照组的注意偏好。

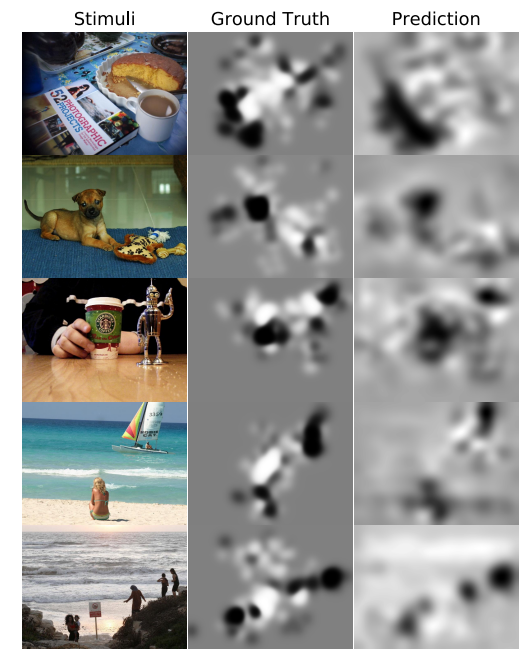


图 7新图像示例上的预测DoF贴图。

1. 结论和今后的工作

临床应用可以显著受益于机器学习的最新进展，以提供更敏感和更具体的诊断。为此，我们提出了一个基于DNN的模型来识别自闭症患者，在自由图像查看中使用他们的眼球跟踪数据。这种数据驱动的方法可以自动学习自闭症的注意力特征，而不依赖于该障碍的任何先验知识。从健康人和自闭症患者的眼动模式的差异中学习辨别特征。实验结果表明，该方法具有较高的灵敏度和选择性。深度神经网络特征和支持向量机分类的显著预测价值可用于支持ASD诊断的临床实践。最后，DNN特征的可视化通过一般和生态相关的自然场景图像提供了对自闭症可能的注意特征的见解，从而进一步探索自闭症的神经病理学。

我们的概念验证模型显示出了有希望的结果，这是首次尝试识别患有自闭症的高功能成年人。为了证明其普遍性，这项工作即将扩展到其他临床人群，例如ASD和其他神经发育障碍的不同亚组。一个更大的眼动数据数据库和不同的主题组也将用于开发和基准测试未来的模型。