**EL-GAN: Embedding Loss Driven Generative Adversarial Networks for Lane Detection**

作者：Mohsen Ghafoorian, Cedric Nugteren, N´ora Baka, Olaf Booij, Michael Hofmann

翻译：蒋明昊，林鑫杰

**摘要**

卷积神经网络已成功应用于诸多语义分割问题。然而，许多问题本质上不是逐像素的分类问题，但是仍经常被表述为语义分割问题。因此，这种不恰当的建模方式需要人们根据特定场景设计特定模型，另外将每个像素的概率分布转换为所需的输出时也需要巨大的计算量。生成对抗网络（GAN）可使得语义分割的输出更加真实、自然，并更好地保留结构信息后期处理的复杂度也更低。在本文中，我们提出EL-GAN：一个缓解嵌入丢失问题的GAN框架。使用EL-GAN，我们可以根据学习到的嵌入表示同时区分标注数据和预测数据。因为模型可以同时得到“错误”和“正确”的预测结果，模型可以学习到区分度更高的有效信息，从而模型训练过程也更加稳定。TuSimple车道标记任务评测结果表明，我们提出的方法是可行的，它可以克服将其作为语义分割问题而带来的固有缺点。与传统方法相比，EL-GAN模型的输出结果不仅与实际结果更贴近，后期处理也相对更简单，并以96％的准确度阈值超越了相同领域竞争对手。

**1 介绍**

卷积神经网络（CNN）已成功用于解决各种计算机视觉问题。相关案例包括对自动驾驶[1,2,3]的道路场景理解和医学成像[4,5,6,7,8,9]。这种网络输出每个像素所属的类别概率。然而在许多情况下，问题本身并不是逐像素的分类任务，换句话说预测结果需要保留某些结构化信息，而这些结构化信息在逐像素的分类方法中并没有得到足够的反映。例如，如果手头的任务是检测图像中的单个直线，逐像素级别的模型并不能检测出诸如粗细程度、弯曲程度、直线是否唯一这类高级特征。这背后的根本原因是逐像素的建模方法把各个像素点的预测看成是相互独立的，它忽略了像素与像素之间潜在的结构化信息。为了获取这些高层次特征常规方法需要额外的后期处理步骤。这些步骤通常包括使用条件随机场（CRF）[10]，单独训练一个网络[2]或是人工设计某种非网络学习的特定算法[11]。这些方法的缺点是设计成本高，超参数多，不具一般性，并且可能仍然无法捕获所需特征。例如，CRF需要单独训练，而且因为它的复杂度很高，是实际使用过程中要么仅捕获局部特征，要么在消耗大量的计算资料来预测长序列。

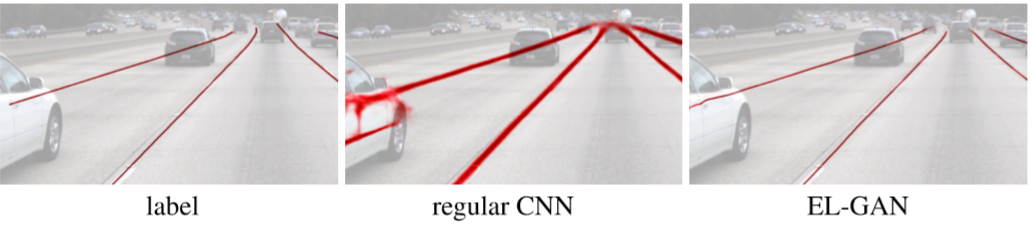


图1：使用EL-GAN进行车道标记分割的示意图：示例地面实况标签（左），基于[17]（中）的传统分割网络对应的原始预测，以及EL-GAN的预测（右））。请注意EL-GAN在确定处和连接处和标注数据的细线是吻合的。

生成对抗网络（GAN）[12]是一种在原来不强调结构化信息的语义分割问题中提取结构化信息的一种很有潜力的工具，它可以自动学习所需的结构化属性。GAN通过在极小极大博弈中以交替方式训练两个网络来工作：训练生成器以产生预测结果，同时训练鉴别器将真实标签（“正确”）数据和预测数据（“假的”）区分开来。GAN也已应用于语义分割问题，使用某种逐像素损失来解决上述问题[1]。通常是生成器先产生语义分割图，然后鉴别器观察实际标注数据和预测数据。正如文献[16]中发现的那样，这种方法存在很明显的不足：鉴别器的二分类预测不能提供稳定的梯度反馈来训练网络。

一般来说，GAN中的鉴别器无监督训练的，它交替地（例如[1]）观察“真”或“假”数据。然而，在语义分割问题中我们通常可以得到对应某预测数据的实际标注数据。而我们所做工作就是想方设法把标注数据和预测数据输入到网络里面，这样鉴别器可以获得更有效的反馈，使得网络训练朝着生成更真实的数据方向发展。换句话说，这种方式可以促使鉴别器能学习到一个带监督的损失函数。

在本文中，我们提出一种在语义分割问题中强调结构化信息的网络结构。具体地说，我们提出EL-GAN（Embedding loss GAN）网络，其中鉴别器将源数据、预测图和实际标注数据作为输入，通过训练最小化预测结果和真实标签之间的差异。EL-GAN的梯度反馈更加有效和训练更家稳定，这使我们能够成功基于GAN架构的训练语义分割网络。因此，我们的模型在不需要额外的问题特定损失项和复杂的后处理步骤的情况下就能使预测结果拥有和训练数据相似的结构化信息。图1中实际数据、常规方法得到的预测数据和EL-GAN得到的预测数据之间的对比就很好的显示了这一点。本文的贡献是：

* 针对语义切割的问题我们提出了一种可以强调结构信息的新型网络结构，它是基于GAN网络架构的，特点在于同时将实际标注数据和预测数据同时输入给鉴别器。这种新网络结构便是我们提出的EL-GAN，使用预测数据和真实标注数据的嵌入表示之间的*L2*损失作为损失函数。
* 我们表明，与普通的对抗损失相比，嵌入损失基本上使训练过程更加稳定并使得梯度反向传播更加有效。与传统的分割网络相比，这不需要额外的损失项或复杂的后期处理工作，但是依然可以产生接近真实数据的预测结果。
* 我们证明了EL-GAN在自动驾驶应用中的实用性，因为该方法是通用的，所以也可以应用于其他语义分段问题。我们在TuSimple车道标记检测数据集对模型进行评测，评测结果表明EL-GAN产生的结果与实际数据更相似。

**2 相关工作**

**语义分割下的特征保持(Quality Preserving Semantic Segmentation)。**有学者提出添加特定损失项[4,9]或使用成对或更高阶项CRF[10,18,19]来强制神经网络保持来某些诸如平滑度，拓扑和局部一致性等特征的方法。但是这些方法大多只能保持较低级别的特征，并且他们的计算成本也相当高。人工设计保持某种特征的损失项是很困难的，难点主要有两个：寻找合适的特征以及相应的可微损失项。

**语义分割的对抗训练(Adversarial Training for Semantic Segmention)。**GANs[12]的主要基本思想是通过鉴别器网络的极小极大博弈训练学习到目标样本的概率分布。Luc等人[1]采用对抗性训练进行分割，以保留更高级别的语义特征。在他们的模型中，鉴别器会把标签和预测分布之间的差异反馈给生成器。这种模型与其他模型的不同之处在于，损失项是网络自动学习得到，不是通过人工设计得到。后来人们将这种机制应用于图像到图像的转换[20]，医学图像分析[5,6,7,8,16,21,22,23]和其他分割任务[24]。然而这种对抗训练并没有使用到图像和标签之间的配对信息。基于此，有学者[25,26]提出以半监督方式使用GAN，并假设未标记数据与标记数据来自相同的分布的方法。EL-GAN模型的原理与此类似，但是我们依认为[25,26]的方法没有保留配对信息，于是我们显式使用配对信息，强制让模型重视这些配对信息。另一项工作[16]提出使用*L1*损失项的医学图像分割GAN网络，但这个模型并没有很强的可解释性，[16]也没有做模型简化研究。与之相比，EL-GAN在鉴别器接收的输入以及用于训练的损失项等方面都有所不同。Hwang等人[27]提出使用生成对抗网络来完成实际图像和预测图像之间的结构匹配的方法，与我们的工作相反，[27]不对鉴别器的输入图像进行限制，也不使用像素级损失。因此，他们得到的鉴别器表示需要保持低水平以确保分段符合低级细节。此外，我们也对EL-GAN进行了简化分析，使得我们的方法更具可解释性。

**感知损失(Perceptual Loss)。**最近的一些工作[28,29,30]，特别是针对图像超分辨率的工作，都普遍认为像素级目标损失通常不足以生成图像的高层语义。因此，他们建议从给定层的单独网络表示中捕获图像的高级表示。在图像超分辨率问题中，通常可获得给定低分辨率图像对应的高分辨率图像。因此，预测图像和真实图像的高层表示之间的差异损失（例如*L2*）被认为是有效的损失项。类似的，在高层嵌入空间中使用标注数据和预测数据之间的差异作为损失项启发了我们。

**车道标记检测(Lane Marking Detection)。**由于我们的工作评估侧重于车道标记检测，我们还讨论了针对此问题的其他相关方法，同时我们建议读者参阅在这个领域的最新调查报告[11]。由Pan等[3]提出的车道标记检测方法是较为成功的一个例子。该方法的核心思想是使用CNN网络配上一层人工设计的后期处理层。Lee等人[31]使用消失点标签来学习到车道标记检测中结构化特征。另一个例子是Neven等[2]的工作，他们使用常规分割网络来获得车道标记预测图。然后，他们训练第二个网络来执行约束变换，之后使用曲线拟合来获得最终结果。针对Tusimple挑战任务，我们将我们的工作与上面的研究[2,3]进行了详细的比较，比较结果在6.1节中展示。

**3 方法**

在本节中，我们将介绍EL-GAN的损失函数和对抗性训练。该方法是通用的，可以应用于各种分割问题。详细的网络架构将在第4节中讨论。

**3.1 基线：语义分割的对抗性训练**

对抗训练可用于确保更高水平的标签，如平滑度，局部一致性等。这是通过使用鉴别器网络来完成的，该网络随着训练的不断深入而学习到所期望属性的损失函数，而不是显式地计算这些属性。从语义分割的对抗性训练中获益的典型方法[1,20]关于生成器制定一个由两个损失项组成的损失函数：一个低层次逐像素级的预测和标签之间的损失项（*L*fit），另一个对抗性损失项用于保留更高级别特征的损失项（*L*adv），给定输入图像：

其中x和y分别是输入图像和相应的标签图,和是生成器和鉴别器网络的参数，表示输入图像x的变换，由生成器网络的参数控制。λ表示对抗性损失项的重要性。 损失项*L*fit通常用逐像素级的分类交叉熵损失表示，其中其中c表示目标类的数量，w和h是图像的宽度和高度。

对抗性损失项*L*adv表示鉴别器拒绝由生成器产生的假预测图像的能力，并且通常用零和鉴别器对生成器生成的图像的预测的二分类交叉熵损失来表达：，其中,且D是鉴别器网络的变换结果。

虽然生成器经过训练来最小化其对抗性损失项，但鉴别器试图通过最小化它自己的损失项使最大化，其中

通过这两个网络的交替训练，鉴别器通过学习真实标签和预测分布之间的差异，进而使生成器提升其预测结果的质量，使得两个分布不可区分。在实践中，与训练正常网络相比，对抗性网络的训练往往更加不稳定。这是因为在极小极大博弈训练中两个网络相互训练，其中每个网络的训练动态影响到另一个网络的训练。鉴别器将生成器生成的图像的接受程度反馈给生成器。通常语义分割对抗训练框架有两个重要问题：

1.预测结果的合理性打分是依赖鉴别器所学习到的对图像特征的表示以及它内部权重编码这些特征的方法。然而有可能鉴别器学习到的这些表示或编码方法根本就不正确，导致生成器梯度下降过程根本就没有朝着改善网络的方向去发展。

2.传统的对抗性损失的损失项没有显式地使用图像和标签之间的配对信息，而这些信息通常是可用于许多监督的语义分割任务。

**3.2 嵌入式损失的对抗性训练**

鉴于上述两个问题，人们可以利用图像/标签配对信息。实现这种想法的一种方式是使用鉴别器将预测/标签图转换为更高层次的嵌入向量，并将对抗性损失定义为嵌入空间内的差异：

我们建议用嵌入损失来表示，定义为嵌入距离（例如*L*2）：

其中表示从用θ参数化的网络D中给定层提取的嵌入向量表示，预测和x作为其输入。

我们将其命名为EL-GAN架构，对抗性损失和相应的梯度是基于标签和预测的嵌入表示之间的差异来计算的。鉴别器最小化其在真实和假分布之间的损失，并且可能学习一组判别嵌入，生成器尝试最小化该嵌入差异。这种生成器的结构图如图2右侧所示。

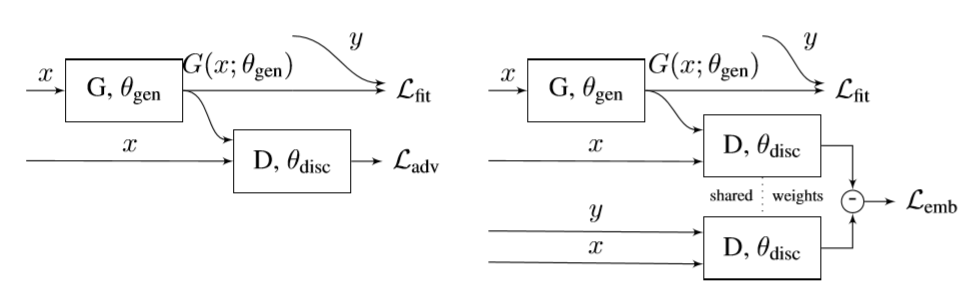


图2：生成器结构示意图：左侧为传统GAN（公式1），右侧使用嵌入损失（公式3和4）

除了上面提到的计算发生器更新的对抗性损失的变化之外，可以选择使用类似的想法重写鉴别器的损失项：

在我们的实证研究中，我们发现使用交叉熵损失来更新鉴别器参数可以得到更好的结果。

**4 实验**

在本节中，我们将详细介绍用于评估方法的数据集和度量标准，然后详细介绍网络细节和训练网络的方法。

**4.1 评估数据集和指标**

我们将评估重点放在自动驾驶应用领域，然而EL-GAN反映的方法是通用的。这项工作的动机之一是能够尽可能地产生与真实标注数据相似的预测结果。~~这对于具有薄结构的TuSimple车道标记检测数据集效果突出，减少了对复杂后期处理的需要。~~

TuSimple车道标记检测数据集1包括3626个带标注的1280×720前向道路图像，这些图像位于圣地亚哥地区的美国高速公路上，分为四个序列，测试图像有2782个。车道标记以折线形式给出：当前车道、左侧车道和右侧车道。折线以每20个像素固定的高度间隔给出。为了生成用于语义分割的标签，我们通过使用具有1像素宽的高斯的平滑插值的离散线条来将这些转换为分割图。这种标签的一个例子在图1的左边以红色显示。

数据集在结果上以与标签相同的格式进行评估，即多条折线。对于我们的评估，我们使用挑战1中定义的官方指标，即准确性，误报率和假阴性率。我们不但在官方测试集上进行了测试，而且还从四个带标注的序列中，选择一个作为验证集进行测试，这个验证集包含409张图像。我们注意到，由于这个验证集的尺寸较小，此验证集上的结果可能并不完全具有代表性。

由于我们的网络仍然输出分割图而不是所需的折线，我们做了一些后期处理。但尽可能让这些处理过程简单化：在二值化之后，我们通过取一系列的非0的y坐标对应的x左边的平均值将每个连通分量转换为单独的折线。我们将此方法称为’basic’。我们还评估了一个'basic ++'版本，该版本会拆分连接的组件，以防它检测到在一个采样位置出现多个非零序列。。

**4.2 网络架构和培训**

在本节中，我们将讨论用于实验的网络和训练的超参数。图3中的高级网络体系结构示意图显示了用于生成器或鉴别器训练的损失项。

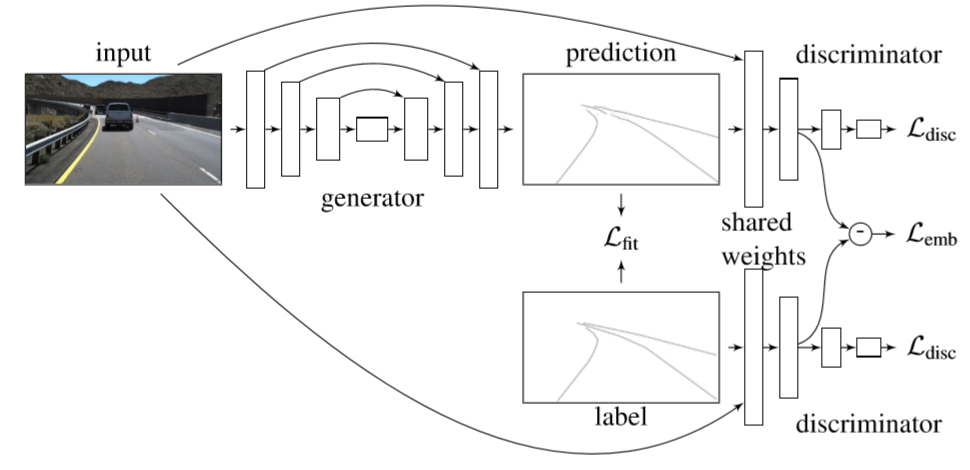


图3：EL-GAN架构概述

对于生成器，我们使用完全卷积U-Net型网络，这种网络具有向下和向上路径以及跳跃连接的特性。特别是我们使用Tiramisu DenseNet架构[17]进行车道标记检测，配置了7个上/下级别，总共64个3×3卷积层。

对于鉴别器，我们使用具有7个块和总共32个3×3卷积层的DenseNet架构[32]，然后是完全卷积分类器[33]。我们使用双头网络为前两个密集块分别处理来自标签或预测的输入图像，之后我们连接特征图。我们在最终卷积层之后提取出嵌入表达，具体的细节在5.2节给出。

我们首先预先训练生成器模型直到收敛，我们也将其用作第5节的非GAN的基线模型。使用批量大小为8，然后我们预先训练一万次迭代的鉴别器，之后在300到200次迭代之间交替训练生成器和鉴别器。使用Adam优化器训练生成器，而使用随机梯度下降SGD来训练鉴别器。我们使用正则化交叉熵损失（等式2）训练鉴别器，同时我们用λ=1的对抗嵌入损失训练生成器（等式3和4）。我们没有进行任何数据扩充，也没有在其他数据上预先训练模型。

**5 实验结果**

在本节中，我们展示使用第4节所讨论的超参数的EL-GAN在TuSimple数据集上得到的测评结果。此外，我们还进行了三项模型简化研究（Ablation study）：评估训练稳定性，探索训练损失的选项，以及改变嵌入损失层的选择。

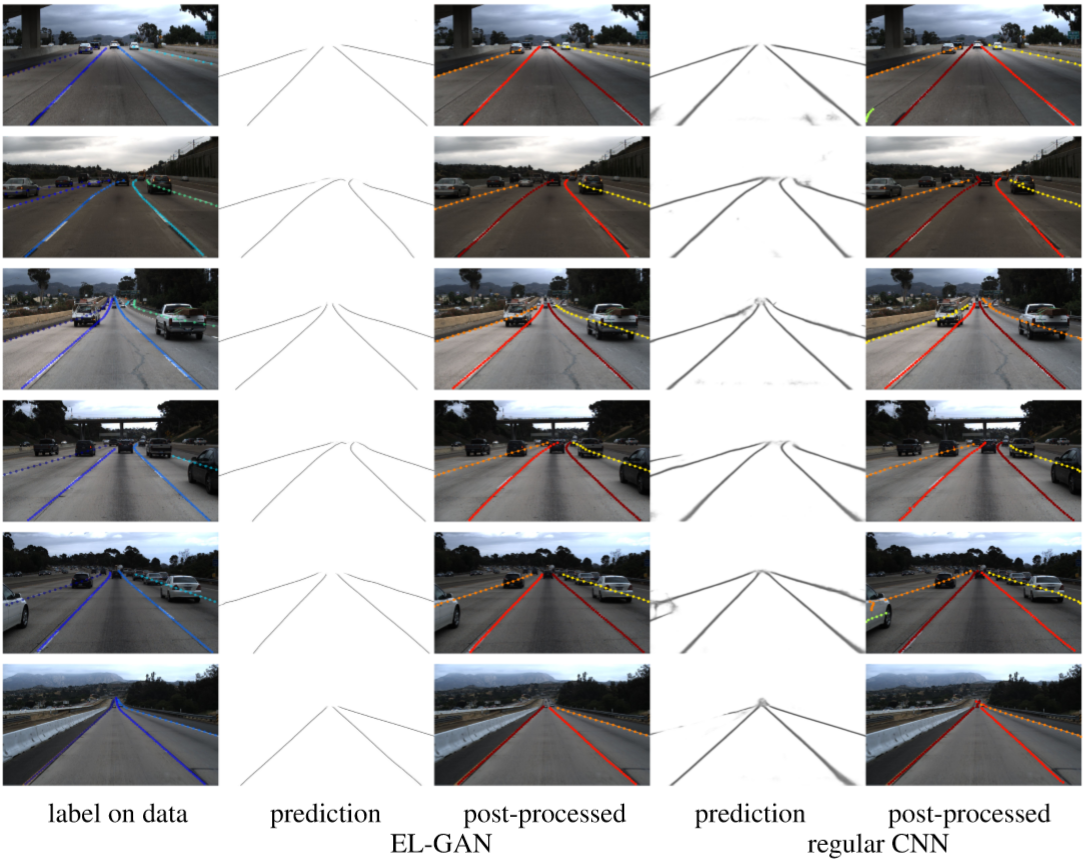


图4：车道标记分段的示例结果：数据顶部的标签（左栏），EL-GAN的预测和最终结果（接下来的两列），以及使用相同后处理的常规CNN基线[17]的结果（右两列）。除了将它们彼此区分之外，线条的颜色没有任何意义。放大时，最好在计算机屏幕上查看详细信息。

表1：TuSimple车道标记验证集的结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 方式 | 后期处理 | 准确率 | FP | FN |
| Baseline(no GAN) | basic | 86.2 | 0.089 | 0.213 |
| Baseline(no GAN) | basic++ | 94.3 | 0.084 | 0.070 |
| EL-GAN | basic | 93.3 | 0.061 | 0.104 |
| EL-GAN | basic++ | **94.9** | **0.059** | **0.067** |

**5.1 TuSimple 车道标记挑战赛**

在本节中，我们将报告TuSimple车道标记检测挑战的结果，并将其与我们的基线和最新技术进行比较。

我们首先使用两种后处理方法评估EL-GAN和验证集上的基线。表1中的结果表明，基本的后期处理方法不适合基线模型，而改进的basic++方法表现更好。尽管如此，EL-GAN仍然优于基线，特别是采用最简单的后期处理方法。

验证集上的一些结果如图4所示，它使用basic++方法比较两种方法的原始预测图和后处理结果。很明显，EL-GAN产生相当薄的和更像标签的输出，噪声更小，使得后期处理更容易。

表2：截至2018年3月14日的TuSimple车道标记挑战排行榜（测试集）

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 等级 | 方法 | 标记名 | 额外数据 | 准确率 | FP | FN |
| #1 | Unpublished | Leonardoli | ? | **96.87** | 0.0442 | 0.0197 |
| #2 | Pan et al.[3] | XingangPan | Yes | 96.53 | 0.0617 | **0.0180** |
| #3 | Unpublished | Aslarry | ? | 96.50 | 0.0851 | 0.0269 |
| #5 | Neven et al.[2] | DavyNeven | No | 96.38 | 0.0780 | 0.0244 |
| #6 | Unpublished | Li | ? | 96.15 | 0.1888 | 0.0365 |
| #14 | Baseline(no GAN) | N/A | No | 94.54 | 0.0733 | 0.0476 |
| #4 | EL-GAN | TomTom EL-GAN | No | 96.39 | **0.0412** | 0.0336 |

表3：不同训练迭代（每10K）的TuSimple验证集准确度统计，比较不同选择的对抗性损失的稳定性

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 损失 | | 准确率统计： | | | |
| 生成器 | 鉴别器 | 均值 | 方差 | 最大值 | 方程式 |
| Cross entropy | Cross entropy | 33.84 | 511.71 | 58.11 | 1 and 2 |
| Cross entropy | Embedding | 0.00 | **0.00** | 0.02 | 1 and 5 |
| Embedding | Cross entropy | 93.97 | 0.459 | 94.65 | 3,4 and 2 |
| Embedding | Embedding | **94.17** | 0.429 | **94.98** | 3,4 and 5 |

此外，我们在整个标记数据集上训练EL-GAN和基线，并使用TuSimple挑战项目的官方测试集上的basic++后期处理进行评估。表2显示了结果，其中包括前6名中的所有方法（据我们所知，其中只有两种是公布的）以及截至2018年3月14日在排行榜上的排名。我们根据准确性排名第4差异小于百分之五，并获得最低的误报率。与基线相比，我们的对抗训练算法在准确度上提高了约2％（误差减少了38％），在私人挑战测试集上将FP降低了55％以上，FN降低了30％。这些改进将基线从第14位提升到第4位。

**5.2模型简化研究**

表3比较了嵌入/交叉熵作为生成器和鉴别器网络训练的对抗性损失项的不同选择的使用。为了比较训练过程的稳定性，我们采集了验证集上准确度的统计数据。图5显示出了验证集F-score平均值和5次训练的标准方差。这些结果表明，使用嵌入损失的生成器可以使GAN训练稳定。我们在使用其他超参数进行训练时也观察到类似的现象。

用于嵌入损失的特征可以在鉴别器中的不同位置处获得。在本节中，我们将探讨三种不同位置的选择：在第3，第5或第7个密集块之后。我们注意到第3个块包含第一个共享卷积层，其中包含输入图像和预测或标签，第7个块包含网络分类器之前的最终卷积结果。TuSimple车道标记检测验证集的结果在表4和图6中给出。从结果中，我们得出结论，特征采集的位置越后，得分越高，预测与标注数据越相似。

**6 讨论**

**6.1 与其他车道标记检测方法的比较**

在表2中，我们在TuSimple车道标记数据集上显示了结果，EL-GAN在排行榜上排名第4。在本节中，我们将我们的方法更详细地与其他两种已发表的方法进行比较：Pan等[3]（排名第2）和Neven等[2]（排名第5）。

Neven等人[2]在他们的工作中强调说，曲线拟合等后期处理技术最好别在网络输出层上进行，而是在俯视角度下进行。他们单独训练一个网络来学习单应性，以找到一个易于曲线拟合的透视变换。在我们的工作中，我们认为可以在不进行曲线拟合的情况下得到类似的效果，从而省略了为达到此目的而进行单独训练网络的步骤。

Pan等[3]使用多分类的方法进行车道标记检测，其中每个车道标记是一个单独的类别。虽然这简化了后期处理，但它在标签创建方面更加复杂，并且使得网络难以解决这个任务：网络模型还应该学会分辨哪个车道是哪个车道，这需要更大的视野并且分辨在车道变换时产生的歧义。相比之下，通过我们的GAN方法，我们可以学习更简单的单类问题，而无需复杂的后期处理来分离单个标记。Pan等人[3]也认为，在进行最终预测之前，诸如车道标记检测之类的问题可以在预测前的空间一致性和信息传递中获益。基于这个原因，他们建议将常规分段网络的输出馈送到特定问题的“Spatial CNN”，其中消息通过不同方向的卷积传递。与EL-GAN相比，这确实导致TuSimple数据集的准确度更高，然而目前还不清楚他们使用了多少训练空间用于CNN，以及实际上他们在一个比TuSimple常规数据集大20倍的非公开数据集上进行训练的具体情况。

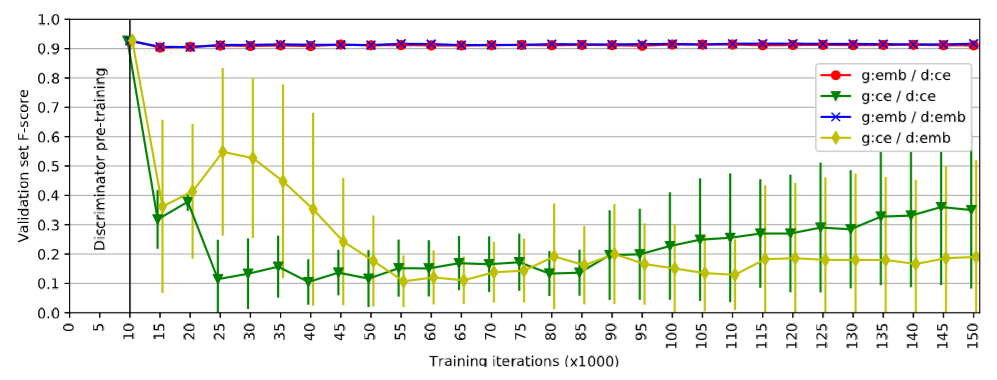


图5：使用不同的对抗性损失项（嵌入/交叉熵）对验证f分数的训练稳定性的比较。对于每种方法，中心点代表平均f分数，每侧的条形图表示标准偏差。应该注意的是，在和情况下，由于不同运行之间的微小变化，标准条不可见。

表4：嵌入提取层的烧蚀研究

|  |  |  |  |
| --- | --- | --- | --- |
| 阻止后嵌入损失# | 准确率(%) | FP | FN |
| Dense block 3 (first block after joining) | 93.91 | 0.1013 | 0.1060 |
| Dense block 5 | 94.01 | 0.0733 | 0.0878 |
| Dense block 7 (before classifier) | 94.94 | 0.0592 | 0.0673 |

**6.2 模型简化研究分析**

正如我们在表3和图5中所示的不同对抗性损失项比较中观察到的那样，使用嵌入损失的生成器可以使得训练更加稳定并防止梯度爆炸。与具有交叉熵损失的通常公式相比，嵌入损失提供了更强的信号，因为它利用现有的基础事实，而不仅仅基于鉴别器的假设及其内部表示。

使用正常的交叉熵损失可能导致模型根本没法训练，这是因为生成器在鉴别器还没有生成对真假图像的理解之前就开始在特征空间进行搜索。相反，使用嵌入损失，这种噪声结果会导致损失项的值变得特别大。准确区分虚假和真实分布的鉴别器也会导致训练不稳定。因此，使用嵌入损失可以使得梯度更好的反馈到生成器，进而得到性能更加优良的生成器。类似地，只在鉴别器使用嵌入损失而不在生成器使用嵌入损失也一样会使训练过程不稳定，主要是因为生成器无法有效的对噪声数据进行惩罚。

在第二项模型简化研究中，表4和图6所示，我们观察到在更深的网络层提取嵌入表达可以获取更好的性能。这可能是由于提取层次越深，这使得生成器能够更好地提取高层次的图像特征。

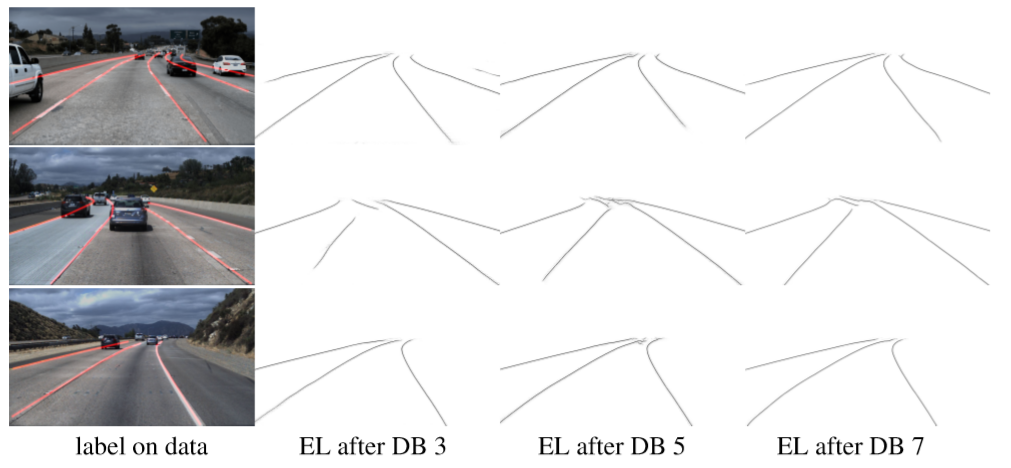


图6：在特定密集块（DB）之后获取嵌入损耗（EL）的比较：数据和标签（左）以及不同设置的预测结果（右三个图像）。 放大时，最好在计算机屏幕上查看详细信息

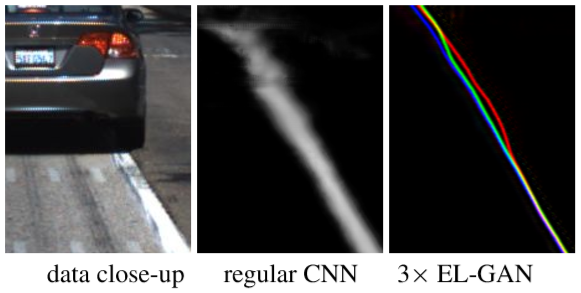


图7：输入数据的示例细节（左），常规语义分段输出（中心），以及使用红色，绿色和蓝色通道显示的相同设置训练的三个不同EL-GAN模型（右）

**6.3 用于语义分割的GAN**

仔细观察常规CNN和EL-GAN之间的比较（图4），我们发现它们的输出性质存在明显差异。非GAN网络根据每个像素每个类别的概率产生概率性输出，而EL-GAN的输出只有一种可能的标签，没有任何不确定性。有人可能会说，缺少不确定性的标识会阻碍了进一步的后期处理。然而，通常应用的后期处理方案的第一步是通过阈值处理或应用最大参数（例如[2,3]）来消除不确定性。另外，普通CNN逐像素独立的概率输出可能隐藏像素间相关性。交叉熵损失使得网络输出不依赖于标签的多样性。

在用于语义分割的EL-GAN和其他GAN中，训练网络在给定输入图像条件下的输出对应同分布的一个标签。图7中示出了一个示例，当车道标记被遮挡或网络变得更不确定时结果的选择过程。虽然这种方式牺牲了某个不确定性，但是它可以使后期处理变得更加容易和准确。对于车道标记检测的任务，我们确实已经表明语义分段不需要输出概率。但是，对于其他应用场景，情况可能并非如此。通过GAN引入不确定性表示的直接方法是在输入中加入一些随机样本并运行多次或使用多个类似的EL-GAN。然后，把这些可能的标签作为输入呈递到后期处理过程。

**7 总结**

在本文中，我们提出、研究和比较了以EL-GAN作为在预测结果中保留标注数据的某些特征的方法。我们发现使用EL-GAN可以对抗训练过程更加稳定。此外，我们在不使用任何额外数据和手工设计的复杂后期处理逻辑的情况下在TuSimple测评数据上取得了目前最好的结果。

**参考**

[1] Luc, P., Couprie, C., Chintala, S., Verbeek, J.: Semantic Segmentation using Adversarial Networks. In: NIPS Workshop on Adversarial Training, Barcelona, Spain (December 2016)

[2] Neven,D.,DeBrabandere,B.,Georgoulis,S.,Proesmans,M.,VanGool,L.: TowardsEnd-toEnd Lane Detection: an Instance Segmentation Approach. ArXiv e-prints (February 2018)

[3] Pan, X., Shi, J., Luo, P., Wang, X., Tang, X.: Spatial As Deep: Spatial CNN for Trafﬁc Scene Understanding. In: AAAI Conference on Artiﬁcial Intelligence. (February 2018)

[4] BenTaieb, A., Hamarneh, G.: Topology Aware Fully Convolutional Networks for Histology Gland Segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2016) 460–468

[5] Dai, W., Doyle, J., Liang, X., Zhang, H., Dong, N., Li, Y., Xing, E.P.: Scan: Structure Correcting Adversarial Network for Chest X-rays Organ Segmentation. arXiv preprint arXiv:1703.08770 (2017)

[6] Huo, Y., Xu, Z., Bao, S., Bermudez, C., Plassard, A.J., Liu, J., Yao, Y., Assad, A., Abramson, R.G., Landman, B.A.: Splenomegaly Segmentation using Global Convolutional Kernels and Conditional Generative Adversarial Networks. Proceedings of SPIE 10574 (2018) 10574 – 10574 – 7

[7] Kohl, S., Bonekamp, D., Schlemmer, H.P., Yaqubi, K., Hohenfellner, M., Hadaschik, B., Radtke, J.P., Maier-Hein, K.: Adversarial Networks for the Detection of Aggressive Prostate Cancer. arXiv preprint arXiv:1702.08014 (2017)

[8] Moeskops, P., Veta, M., Lafarge, M.W., Eppenhof, K.A., Pluim, J.P.: Adversarial Training and Dilated Convolutions for Brain MRI Segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer (2017) 56–64

[9] Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S., de Marvao, A., Dawes, T., O’Regan, D., etal.: AnatomicallyConstrainedNeuralNetworks(ACNN): Application to Cardiac Image Enhancement and Segmentation. IEEE transactions on medical imaging (2017)

[10] Kr¨ahenb¨uhl, P., Koltun, V.: Efﬁcient inference in fully connected crfs with gaussian edge potentials. In: Advances in neural information processing systems. (2011) 109–117

[11] Bar Hillel, A., Lerner, R., Levi, D., Raz, G.: Recent Progress in Road and Lane Detection: a Survey. Machine Vision and Applications25(3) (Apr 2014) 727–745

[12] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Nets. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence,N.D.,Weinberger,K.Q., eds.: AdvancesinNeuralInformationProcessingSystems 27. Curran Associates, Inc. (2014) 2672–2680

[13] Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive Growing of GANs for Improved Quality, Stability, and Variation. ArXiv e-prints (October 2017)

[14] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H.: Generative Adversarial Text-to-Image Synthesis. In: Proceedings of The 33rd International Conference on Machine Learning. (2016)

[15] Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., Metaxas, D.: StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. In: ICCV: International Conference on Computer Vision. (2017)

[16] Xue, Y., Xu, T., Zhang, H., Long, R., Huang, X.: SegAN: Adversarial Network with Multiscale L 1 Loss for Medical Image Segmentation. ArXiv e-prints (June 2017)

[17] Jegou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio, Y.: The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. In: CVPRW: IEEE Conference on Computer Vision and Pattern Recognition Workshops. (July 2017) 1175–1183 [18] Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.: Conditional Random Fields as Recurrent Neural Networks. In: ICCV: International Conference on Computer Vision. (2015) 1529–1537

[19] Schwing, A.G., Urtasun, R.: Fully Connected Deep Structured Networks. ArXiv e-prints (March 2015)

[20] Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image Translation with Conditional Adversarial Networks. In: CVPR: Computer Vision and Pattern Recognition. (2017)

[21] Yang, D., Xu, D., Zhou, S.K., Georgescu, B., Chen, M., Grbic, S., Metaxas, D., Comaniciu, D.: Automatic liver segmentation using an adversarial image-to-image network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2017) 507–515

[22] Li, Z., Wang, Y., Yu, J.: Brain Tumor Segmentation Using an Adversarial Network. In: International MICCAI Brainlesion Workshop, Springer (2017) 123–132

[23] Sadanandan, S.K., Karlsson, J., Whlby, C.: Spheroid segmentation using multiscale deep adversarial networks. In: ICCVW: IEEE International Conference on Computer Vision Workshops. (Oct 2017) 36–41

[24] Nguyen, V., Vicente, T.F.Y., Zhao, M., Hoai, M., Samaras, D.: Shadow detection with conditionalgenerativeadversarialnetworks. In: ICCV:IEEEInternationalConferenceonComputer Vision, IEEE (2017) 4520–4528

[25] Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P., Chen, D.Z.: Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2017) 408–416

[26] Hung, W.C., Tsai, Y.H., Liou, Y.T., Lin, Y.Y., Yang, M.H.: Adversarial Learning for SemiSupervised Semantic Segmentation. ArXiv e-prints (February 2018)

[27] Hwang, J.J., Ke, T.W., Shi, J., Yu, S.X.: Adversarial structure matching loss for image segmentation. arXiv preprint arXiv:1805.07457 (2018)

[28] Sajjadi, M.S., Scholkopf, B., Hirsch, M.: EnhanceNet: Single image super-resolution through automated texture synthesis. In: CVPR: IEEE Conference on Computer Vision and Pattern Recognition. (2017) 4491–4500

[29] Dosovitskiy, A., Brox, T.: Generating Images with Perceptual Similarity Metrics based on Deep Networks. In: NIPS: Advances in Neural Information Processing Systems. (2016) 658– 666

[30] Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual Losses for Real-time Style Transfer and Superresolution. In: ECCV: European Conference on Computer Vision, Springer (2016) 694–711

[31] Lee, S., Kim, J., Yoon, J.S., Shin, S., Bailo, O., Kim, N., Lee, T.H., Hong, H.S., Han, S.H., Kweon, I.S.: Vpgnet: Vanishing point guided network for lane and road marking detection and recognition. In: Computer Vision (ICCV), 2017 IEEE International Conference on, IEEE (2017) 1965–1973

[32] Huang, G., Liu, Z., Weinberger, K.Q.: Densely Connected Convolutional Networks. In: CVPR: Computer Vision and Pattern Recognition. (2017)

[33] Li, C., Wand, M.: Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks. In Leibe, B., Matas, J., Sebe, N., Welling, M., eds.: ECCV: European Conference on Computer Vision, Cham, Springer International Publishing (2016) 702–716

[34] He, K., Zhang, X., Ren, S., Sun, J.: Delving Deep Into Rectiﬁers: Surpassing Human-level Performance on Imagenet Classiﬁcation. In: ICCV: International Conference on Computer Vision. (2015) 1026–1034

[35] Clevert, D.A., Unterthiner, T., Hochreiter, S.: Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). In: ICLR: International Conference on Learning Representations. (2015)

[36] Kingma, D.P., Ba, J.: Adam: A method for Stochastic Optimization. In: ICLR: International Conference on Learning Representations. (2014)

**附录A：网络架构和培训配置**

以下是超参数的详细信息，以确保可重复性。 我们首先列出了用于我们实验的网络架构设置：

* 生成器：Tiramisu DenseNet [17]，向下/向上采样路径中的密集块数量：7，每个密集块中3×3转换层的数量：[1,2,3,4,6,8,8] ，增长率：18，非线性函数：ReLU，初始化：He [34]，dropout率：0.1。
* 判别器：双头DenseNet [32]，连接两个头：第二个密集块之后的串联，密集块的数量：7，每个密集块中3×3个转换层的数量：[1,2,3，4,6,8,8]，增长率：8，非线性函数：ELU [35]，dropout:0，在第7个密集块之后提取嵌入表达。

详细的训练超参数如下：

* 一般：迭代次数：150K，批量大小：8，训练时间表：（300：disc，200：gen）。
* 生成器：优化器:( Adam [36]，动量：0.9），学习速率:(初始值：5e4，decay power：0.99，dacay rate：200），L2正则化标量：1e-4，预训练：100K次迭代。
* 鉴别器：优化器：vanilla SGD，学习速率:(初始值：1e-5，decay power：0.99，decay rate：800），预训练：10K迭代，L2正则化标量：1e-5，对抗性损失λ：1。