

# arXiv论文：自动驾驶中深度学习

是Elektrobit Automotive公司刚刚上传的综述文章（2019/10/17），其安全性的讨论很有启发。

本文目的是研究自动驾驶中深度学习技术的最新技术。首先介绍基于AI的自动驾驶架构、CNN和

RNN、以及DRL范例。这些方法为驾驶场景感知、路径规划、行为决策和运动控制算法奠定基础。

该文研究深度学习方法构建的模块化“感知-规划-执行”流水线以及将传感信息直接映射到转向命令

的端到端系统。此外，设计自动驾驶AI架构遇到的当前挑战，如安全性、训练数据源和计算硬件等也

进行了讨论。该工作有助于深入了解深度学习和自动驾驶AI方法的优越性和局限性，并协助系统的设计选择。

## A Survey of Deep Learning Techniques for Autonomous Driving

---

**Sorin Grigorescu\***

Artificial Intelligence,  
Elektrobit Automotive.

Robotics, Vision and Control Lab,  
Transilvania University of Brasov.  
Brasov, Romania

`Sorin.Grigorescu@elektrobit.com`

**Bogdan Trasnea**

Artificial Intelligence,  
Elektrobit Automotive.

Robotics, Vision and Control Lab,  
Transilvania University of Brasov.  
Brasov, Romania

`Bogdan.Trasnea@elektrobit.com`

**Tiberiu Cocias**

Artificial Intelligence,  
Elektrobit Automotive.

Robotics, Vision and Control Lab,  
Transilvania University of Brasov.  
Brasov, Romania

`Tiberiu.Cocias@elektrobit.com`

**Gigel Macesanu**

Artificial Intelligence,  
Elektrobit Automotive.

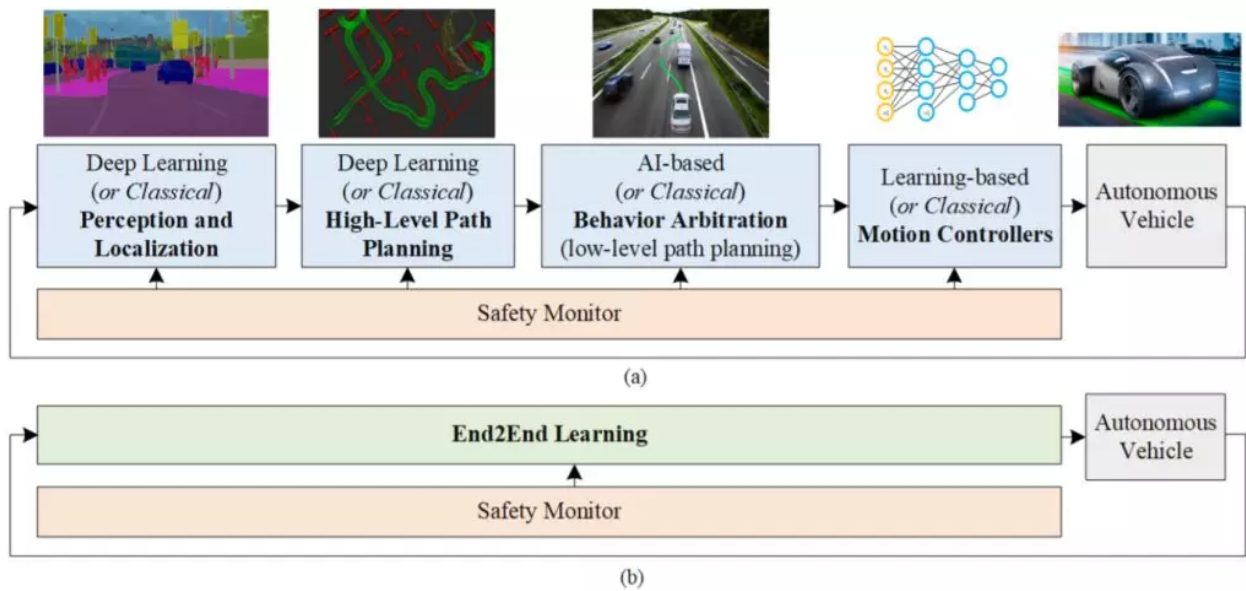
Robotics, Vision and Control Lab,  
Transilvania University of Brasov.  
Brasov, Romania

`Gigel.Macesanu@elektrobit.com`

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Deep Learning based Decision-Making Architectures used in Self-Driving Cars</b>	<b>3</b>
<b>3</b>	<b>Overview of Deep Learning Technologies</b>	<b>4</b>
3.1	Deep Convolutional Neural Networks . . . . .	5
3.2	Recurrent Neural Networks . . . . .	6
3.3	Deep Reinforcement Learning . . . . .	7
<b>4</b>	<b>Deep Learning for Driving Scene Perception and Localization</b>	<b>10</b>
4.1	Sensing Hardware: Camera vs. LiDAR Debate . . . . .	10
4.2	Driving Scene Understanding . . . . .	11
4.2.1	Bounding-Box-Like Object Detectors . . . . .	12
4.2.2	Semantic and Instance Segmentation . . . . .	12
4.2.3	Localization . . . . .	14
4.3	Perception using Occupancy Maps . . . . .	15
<b>5</b>	<b>Deep Learning for Path Planning and Behavior Arbitration</b>	<b>16</b>
<b>6</b>	<b>Motion Controllers for AI-based Self-Driving Cars</b>	<b>16</b>
6.1	Learning Controllers . . . . .	17
6.2	End2End Learning Control . . . . .	18
<b>7</b>	<b>Safety of Deep Learning in Autonomous Driving</b>	<b>20</b>
<b>8</b>	<b>Data Sources for Training Autonomous Driving Systems</b>	<b>23</b>
<b>9</b>	<b>Computational Hardware and Deployment</b>	<b>26</b>
<b>10</b>	<b>Discussion and Conclusions</b>	<b>27</b>
10.1	Final Notes . . . . .	29

如图所示，基于深度学习的自动驾驶汽车。该体系结构既可以实现为串行“感知-规划-执行”流水线（a），也可以实现为端到端系统（b）。在串行流水线中，可以使用AI和深度学习或基于经典的非机器学习方法来设计组件。端到端学习系统主要基于深度学习。这两种方案，通常都设计一个安全监视器来确保每个模块的安全。



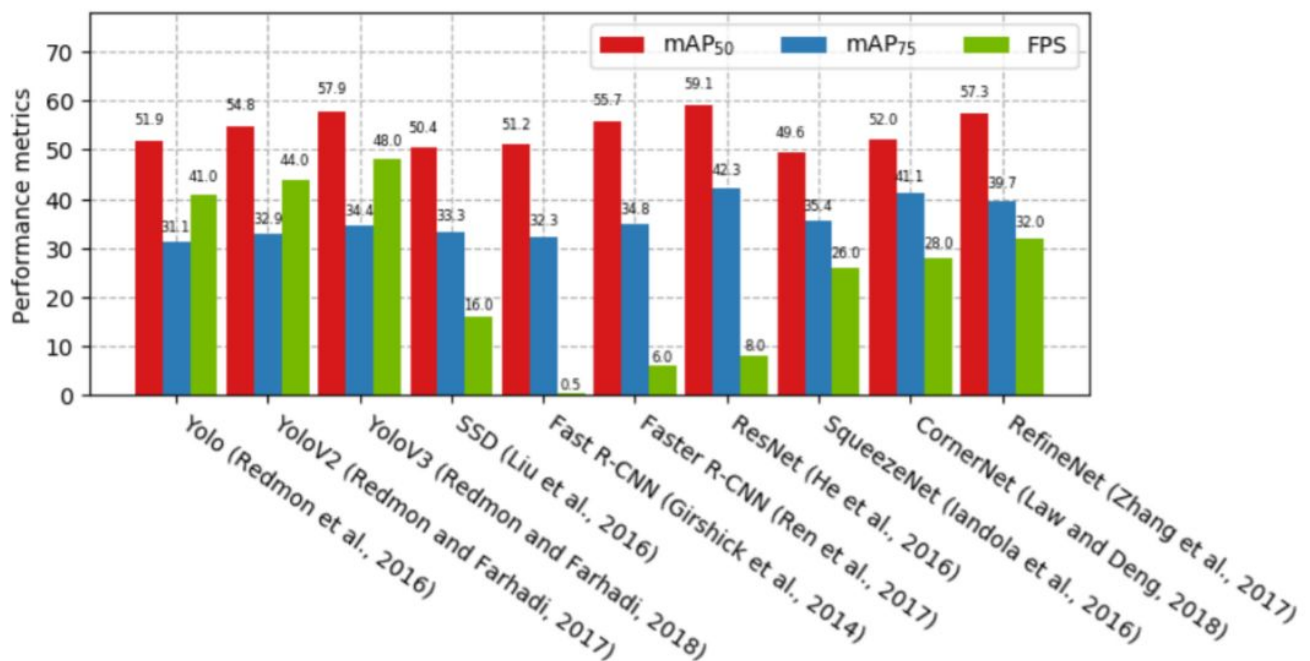
关于深度学习的理论部分跳过。

## 感知和定位

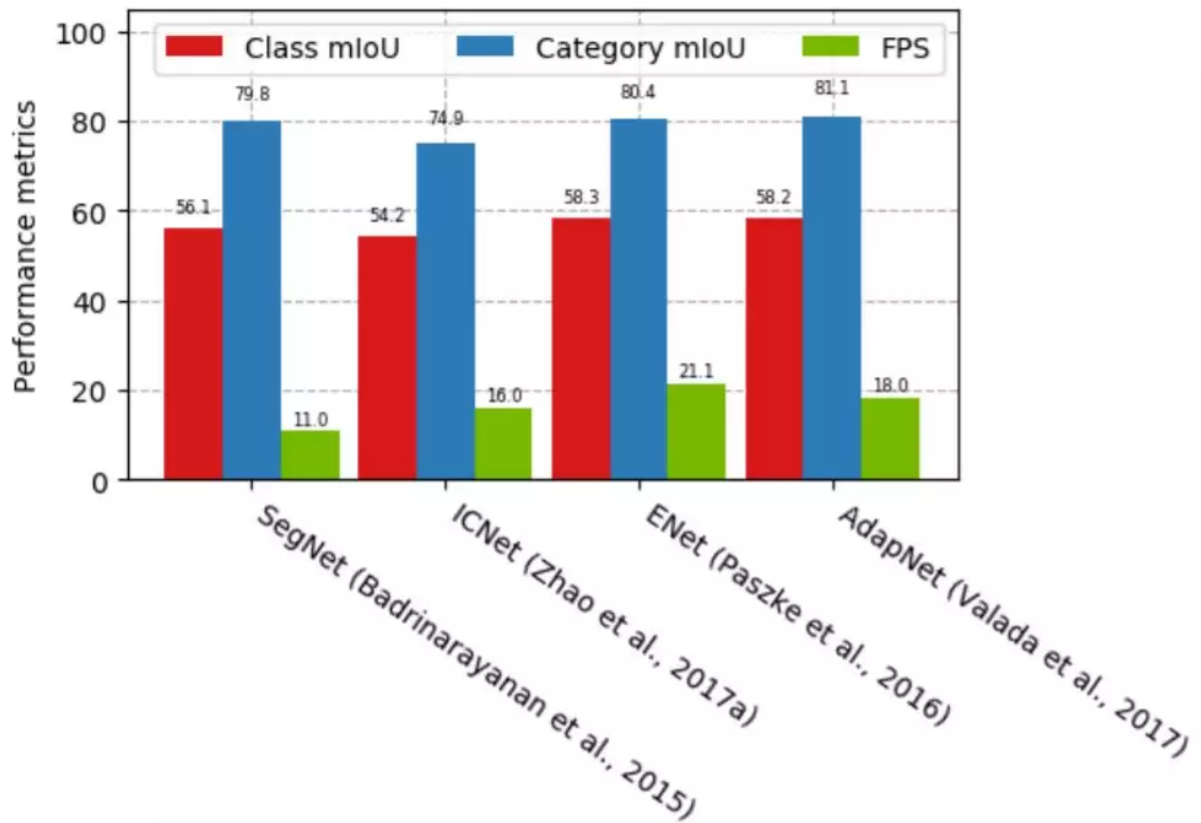
传感器硬件：激光雷达和摄像头之争

场景理解：边框目标检测，语义和实例分割，定位

如图是目标检测方法性能比较。



下图是语义分割方法性能比较。



于占有图 ( occupancy map ) 的感知

下图是一些占有图例子。



## 路径规划和行为决策

模拟学习 ( Imitation learning ) 和深度强化学习 ( DRL )

## 运动控制器

学习控制器 ( Learning controller ) : 迭代学习控制 ( ILC ) 和模型预测控制 ( MPC )

端到端 ( E2E ) 控制器 : DNN , DRL



## 安全性

推理深度学习技术安全性的要求有以下几点：

- 了解可能发生故障的影响；
- 了解更广泛系统的上下文；
- 定义有关系统上下文和可能使用环境的假设；
- 定义安全行为的方法，包括非功能性约束。

**不管各种经验定义和安全性解释如何，在安全关键系统中深度学习组件仍然是一个悬而未决的问题。道路车辆功能安全ISO 26262标准提供了确保安全的全面要求，但并未解决基于深度学习软件的独特特征。**

标准ISO 26262建议用危害分析和风险评估（HARA，Hazard Analysis and Risk Assessment）方法识别系统中的危害事件并指定减轻危害的安全目标。该标准有10个部分。重点是第6部分：软件级别的产品开发，遵循众所周知的V-工程模型标准。汽车安全完整性等级（ASIL，Automotive Safety Integrity Level）是指ISO 26262中某个汽车系统项目（例如子系统）定义的风险方案分类。

ASIL代表降低风险所需的严格程度（例如，测试技术，所需文档的类型等），其中ASIL D代表最高风险，ASIL A代表最低风险。如果一个元素分配给质量管理（QM，Quality Management），那么它不需要安全管理。给定危害评估的ASIL首先分配给旨在解决该危害的安全目标，然后该目标得出的安全要求才接受该目标。

根据ISO26226，危害（Hazard）被定义为“由故障（malfunctioning）行为引起的潜在伤害源，伤害指人身伤害或对人体健康的损害”。但是，**深度学习组件可能会造成新的危险类型。**

由于复杂性，深度学习组件可能会以特有的方式失败。例如，在深度强化学习系统中，奖励函数的错误可能会对训练后的模型产生负面影响。这种情况下，自动驾驶汽车会发现，某些传感器漏洞（sensor vulnerabilities）让它看不到距离其他汽车有多近，从而避免了因距离太近而被惩罚。尽管此类危险可能是深度强化学习组件所特有的，但故障仍可以追踪到，因此适合ISO 26262的现有准则。

一个分析深度学习组件安全性的关键要求是，检查其结果的直接人工成本是否超过某些危害严重性阈值。从人看，不希望有的结果确实是有危害的，并且其作用几乎是实时的。这些结果可以归类为安全问题。深度学习决策的成本与包含明确损失函数L的优化目标有关，即训练误差函数。

实际上，机器学习系统仅遇到有限数量的测试样本，而实际操作风险是测试集的经验量化值。即使算法假设是风险最优的，操作风险也可能比小基数（cardinality）测试集的实际风险大得多。测试集实例引起的不确定性，可能对单个测试样本具有很大的安全隐患。

编程组件的故障和失败（例如采用算法求解问题的组件）与深度学习组件完全不同。深度学习组件的特定故障可能由不可靠或噪声传感器信号（恶劣天气下的视频信号、吸收建筑材料的雷达信号和GPS数据等）、神经网络拓扑结构、学习算法、训练集或意外引起的环境变化（例如未知的驾驶场景或道路事故）造成。自动驾驶车辆必须具有故障安全（fail-safe）机制，通常由称为安全监控器（Safety Monitors）实现。一旦检测到故障，它们必须停止自主控制软件。

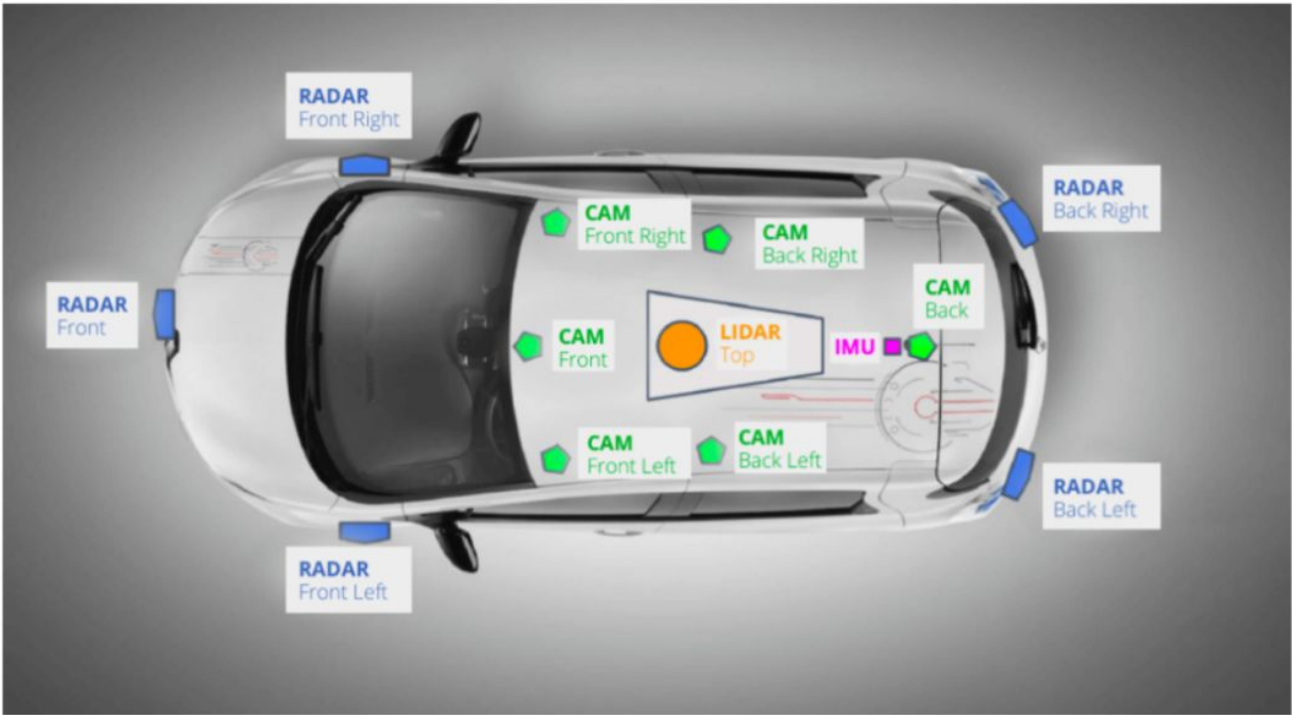
训练集在深度学习组件的安全性中起关键作用。ISO 26262标准规定，应充分规定组件性能（component behavior），并根据其规格对每个修正进行验证。深度学习系统中用训练集代替规范，这违反了该假设。目前尚不清楚如何确保相应的危害始终得到减轻。训练过程不是验证过程，因为训练模型取决于训练集。

应当制定详细的要求并能追踪到危害，这样可以指定如何获得训练集、验证集和测试集。随后，按照这个规范收集的数据得到验证。此外，用某些规范（例如，车辆不能超过3米的事实）来丢弃误报（false positive detections）。

对于将安全视为关键的系统，即使其完整的安全性保证仍然是一个悬而未决的问题，机器学习和深度学习技术还是变得要有效和可靠。**汽车行业当前的标准和法规无法被完全映射到此类系统，因此需要开发针对深度学习的新安全标准。**

### 数据源

现实世界的的数据，是培训和测试自动驾驶组件的关键要求。组件开发阶段需要大量数据，需要在公共道路上收集数据。为了获得对驾驶场景的全面描述，数据收集的车辆配备各种传感器，例如雷达、激光雷达、GPS、摄像机、惯性测量单元（IMU）和超声波等，如图所示。



如图是一些公开的自动驾驶数据集：

Dataset	Problem Space	Sensor setup	Size	Location	Traffic condition	License
NuScenes (Caesar et al., 2019)	3D tracking, 3D object detection	Radar, Lidar, EgoData, GPS, IMU, Camera	345 GB (1000 scenes, clips of 20s)	Boston, Singapore	Urban	CC BY-NC-SA 3.0
AMUSE (Koschorrek et al., 2013)	SLAM	Omnidirectional camera, IMU, EgoData, GPS	1 TB (7 clips)	Los Angeles	Urban	CC BY-NC-ND 3.0
Ford (Pandey et al., 2011)	3D tracking, 3D object detection	Omnidirectional camera, IMU, Lidar, GPS	100 GB	Michigan	Urban	Not specified
KITTI (Geiger et al., 2013)	3D tracking, 3D object detection, SLAM	Monocular cameras, IMU Lidar, GPS	180 GB	Karlsruhe	Urban Rural	CC BY-NC-SA 3.0
Udacity (Udacity, 2018)	3D tracking, 3D object detection	Monocular cameras, IMU, Lidar, GPS, EgoData	220 GB	Mountain View	Rural	MIT
Cityscapes (Cityscapes, 2018)	Semantic understanding	Color stereo cameras	63 GB (5 clips)	Darmstadt, Zurich, Strasbourg	Urban	CC BY-NC-SA 3.0
Oxford (Maddern et al., 2017)	3D tracking, 3D object detection, SLAM	Stereo and monocular cameras, GPS Lidar, IMU	23 TB (133 clips)	Oxford	Urban, Highway	CC BY-NC-SA 3.0
CamVid (Brostow et al., 2009)	Object detection, Segmentation	Monocular color camera	8 GB (4 clips)	Cambridge	Urban	N/A
Daimler pedestrian (Flohr and Gavrila, 2013)	Pedestrian detection, Classification, Segmentation, Path prediction	Stereo and monocular cameras	91 GB (8 clips)	Amsterdam, Beijing	Urban	N/A
Caltech (Dollar et al., 2009)	Tracking, Segmentation, Object detection	Monocular camera	11 GB	Los Angeles (USA)	Urban	N/A

## 计算硬件

Nvidia的Drive PX , Renesas的R-Car。