

---

# End to End Learning for Self-Driving Cars

---

**Mariusz Bojarski**  
NVIDIA Corporation  
Holmdel, NJ 07735

**Davide Del Testa**  
NVIDIA Corporation  
Holmdel, NJ 07735

**Daniel Dworakowski**  
NVIDIA Corporation  
Holmdel, NJ 07735

**Bernhard Firner**  
NVIDIA Corporation  
Holmdel, NJ 07735

**Beat Flepp**  
NVIDIA Corporation  
Holmdel, NJ 07735

**Prasoon Goyal**  
NVIDIA Corporation  
Holmdel, NJ 07735

**Lawrence D. Jackel**  
NVIDIA Corporation  
Holmdel, NJ 07735

**Mathew Monfort**  
NVIDIA Corporation  
Holmdel, NJ 07735

**Urs Muller**  
NVIDIA Corporation  
Holmdel, NJ 07735

**Jiakai Zhang**  
NVIDIA Corporation  
Holmdel, NJ 07735

**Xin Zhang**  
NVIDIA Corporation  
Holmdel, NJ 07735

**Jake Zhao**  
NVIDIA Corporation  
Holmdel, NJ 07735

**Karol Zieba**  
NVIDIA Corporation  
Holmdel, NJ 07735

## Abstract

We trained a convolutional neural network (CNN) to map raw pixels from a single front-facing camera directly to steering commands. This end-to-end approach proved surprisingly powerful. With minimum training data from humans the system learns to drive in traffic on local roads with or without lane markings and on highways. It also operates in areas with unclear visual guidance such as in parking lots and on unpaved roads.

The system automatically learns internal representations of the necessary processing steps such as detecting useful road features with only the human steering angle as the training signal. We never explicitly trained it to detect, for example, the outline of roads.

Compared to explicit decomposition of the problem, such as lane marking detection, path planning, and control, our end-to-end system optimizes all processing steps simultaneously. We argue that this will eventually lead to better performance and smaller systems. Better performance will result because the internal components self-optimize to maximize overall system performance, instead of optimizing human-selected intermediate criteria, e. g., lane detection. Such criteria understandably are selected for ease of human interpretation which doesn't automatically guarantee maximum system performance. Smaller networks are possible because the system learns to solve the problem with the minimal number of processing steps.

We used an NVIDIA DevBox and Torch 7 for training and an NVIDIA DRIVE™ PX self-driving car computer also running Torch 7 for determining where to drive. The system operates at 30 frames per second (FPS).