HUAWEI
EUROPEAN RESEARCH CENTER

Real-time End-to-End Lane ID Estimation Using Recurrent Networks

Dr. Ibrahim Halfaoui

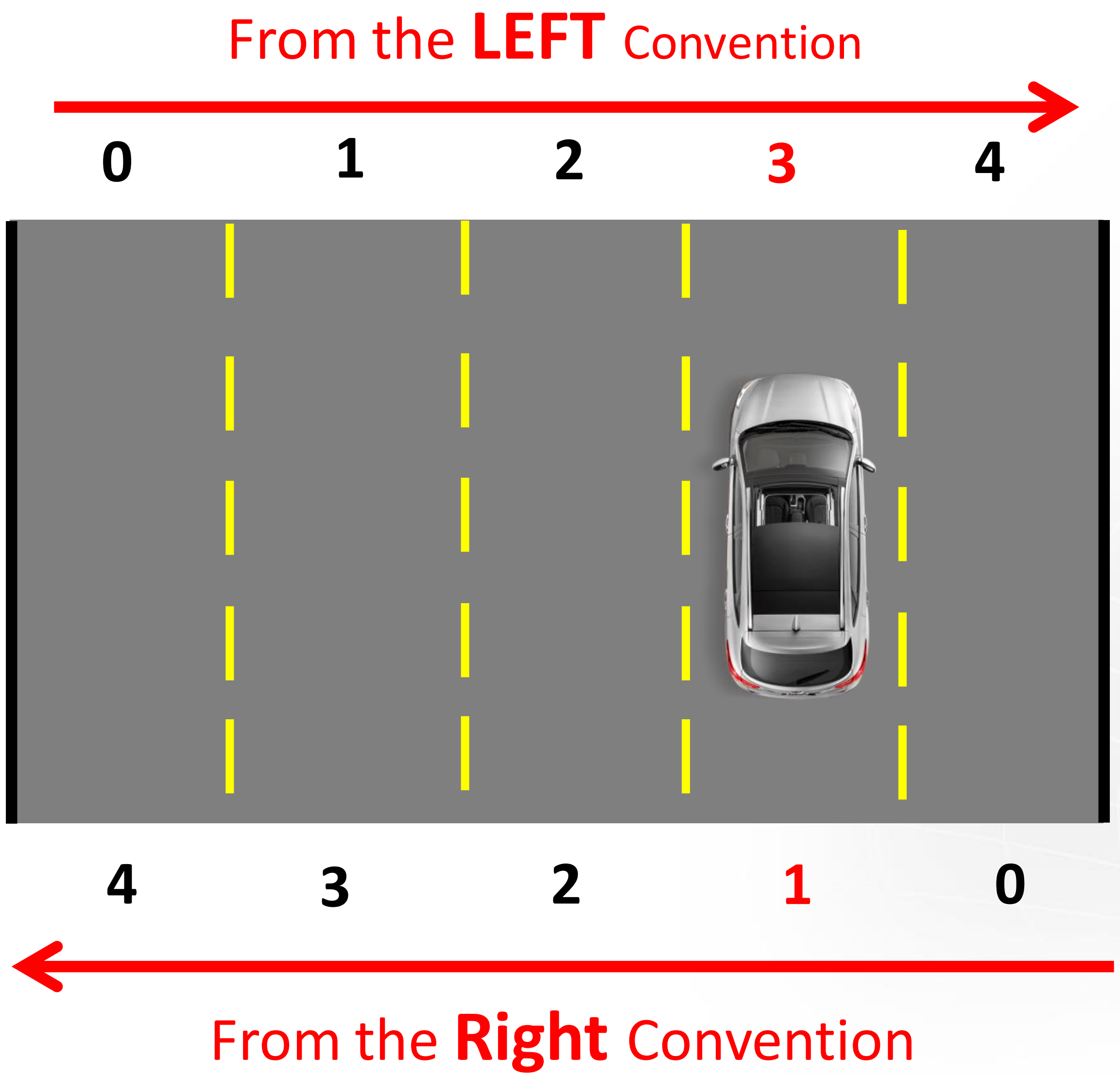
Lane ID Estimation – Task Discerption

- **Task description**: Given a visual representation of a driving scene (image), **the car should be able to identify which lane it is driving on.**
- The identification is done using a number/Identifier of the lane (Lane ID estimation).

End-to-end supervised deep learning approach

Lane ID Estimation – Lane Convention

- For our end-to-end approach, we are using a **Dual ID Convention:**



In this case, the car is on:

- Lane ID $\boldsymbol{\delta l} = \mathbf{3}$ from **left**.
- Lane ID $\boldsymbol{\delta r} = \mathbf{1}$ from **right**.
- C = 5 is total lane count.

$$\boldsymbol{\delta l + \delta r + 1 = C}$$

Lane ID Estimation – Dataset

- **Training** and **Testing** datasets are recorded in Shanghai ➔ Images + Corresponding labels

LANE ID Estimation Dataset

- **5** delivered and processed batches of data recorded between June 2018 and February 2019 (different conditions).

Training Set

- Combined as following:
  - **Batch 1**: 84 Sequences.
  - **Batch 3**: 30 Sequences.
  - **Batch 4:** 87 Sequences.
  - **Batch 5**: 43 Sequences.

244 Sequences
More than 600K images
(~**3 TB**)

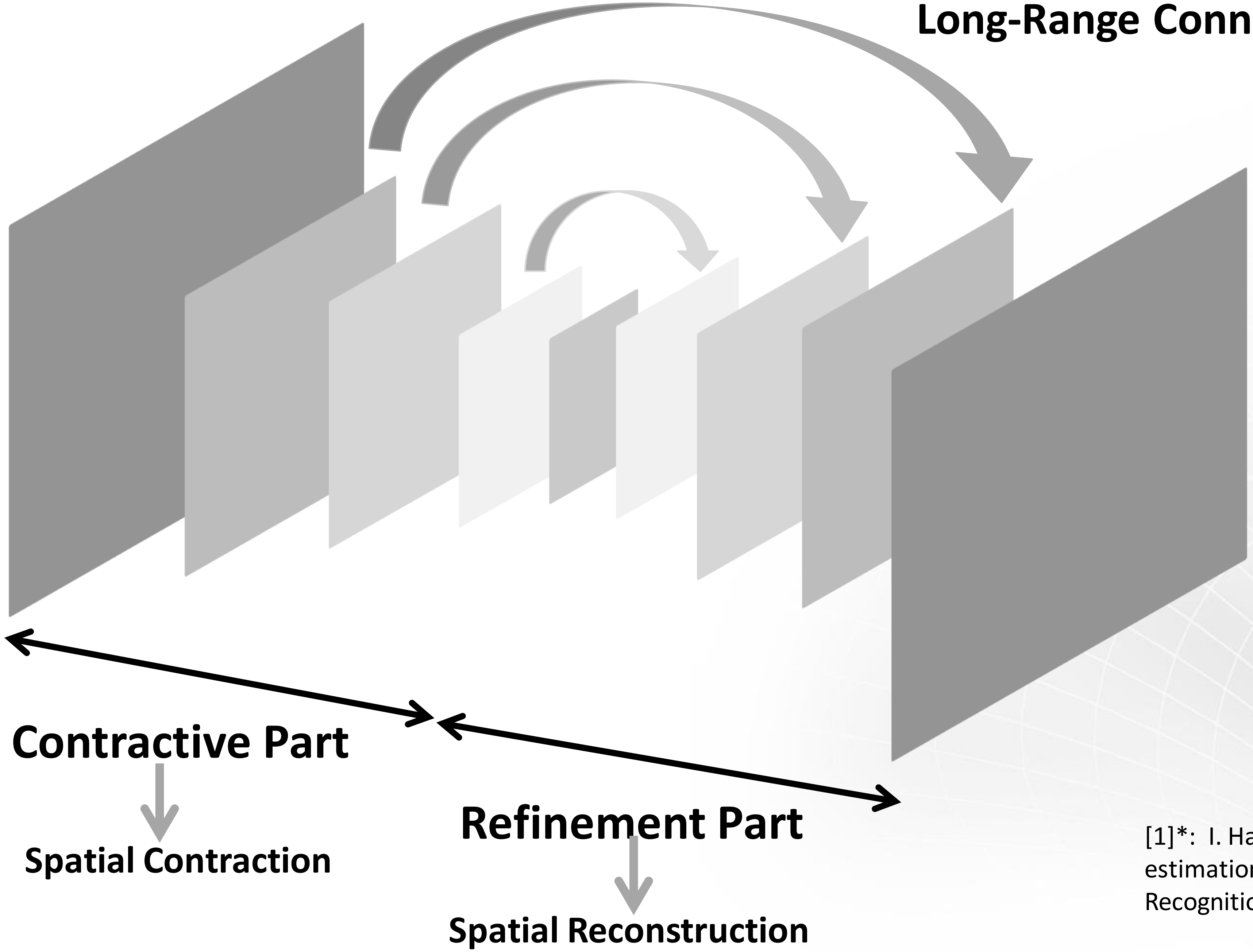
- Composed of:
  - **Day/Afternoon/Noon/Night** sequences (9 night seq.)
  - Good/Bad weather conditions.
  - Training is limited to **up to N = 8 lanes.**

Testing Set

- Combined as following:
  - **Batch 1**: 49 Sequences.
  - **Batch 3**: 27 Sequence (**Night**).
  - **Batch 4:** 51 Sequences.
  - **Batch 5**: 36 Sequences.

163 Sequences
More than 400K
images

Lane ID Estimation – "Moka"-Style Architecture [1]*



- ✓ Per-pixel Prediction
- ✓ Efficient Integration of Global Properties
- ✓ Computational Efficiency

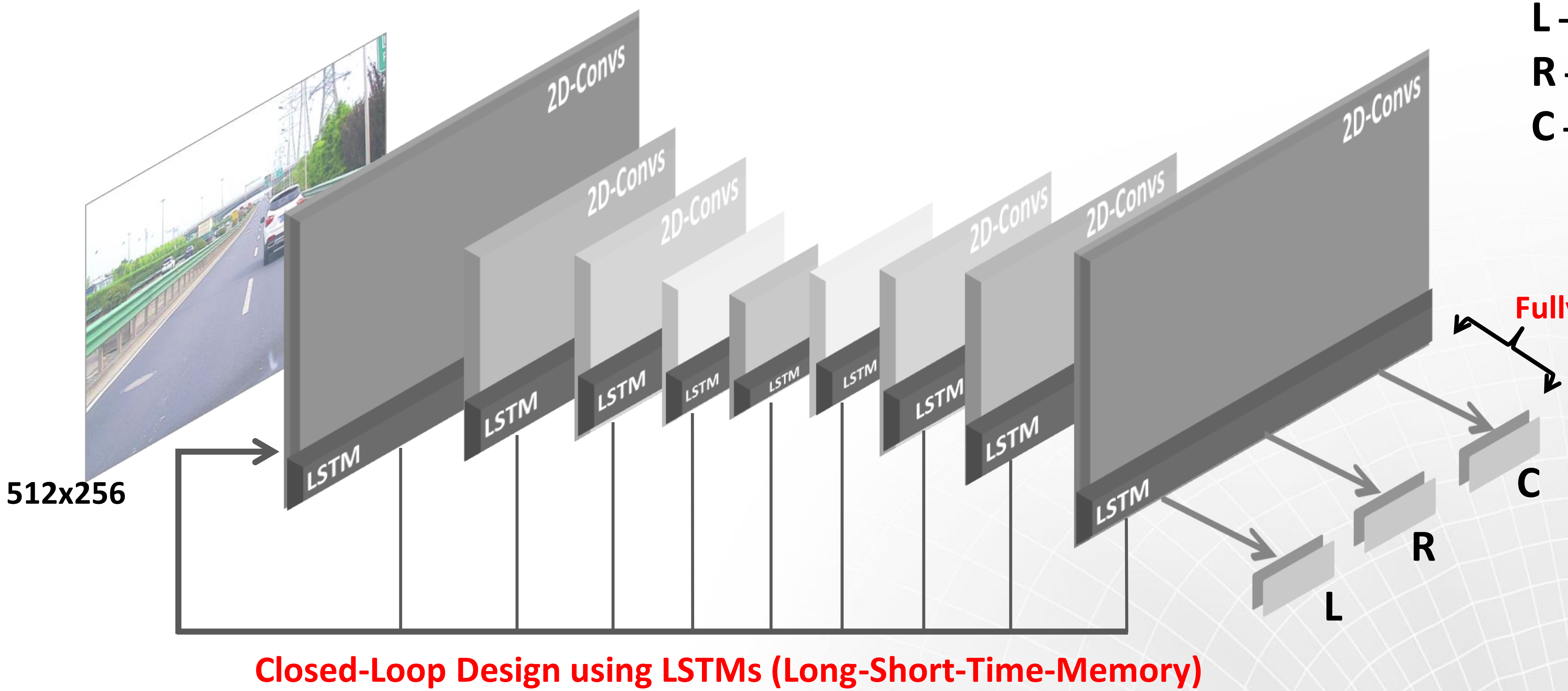
[1]*: I. Halfaoui, F. Bouzaraa, and O. Urfalioglu, "Cnn-based initial background estimation," in 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016, pp. 101–106.

Lane ID Estimation – Moka-LSTM Architecture

L – Left ID estimate

R – Right ID estimate

C – Lane count estimate

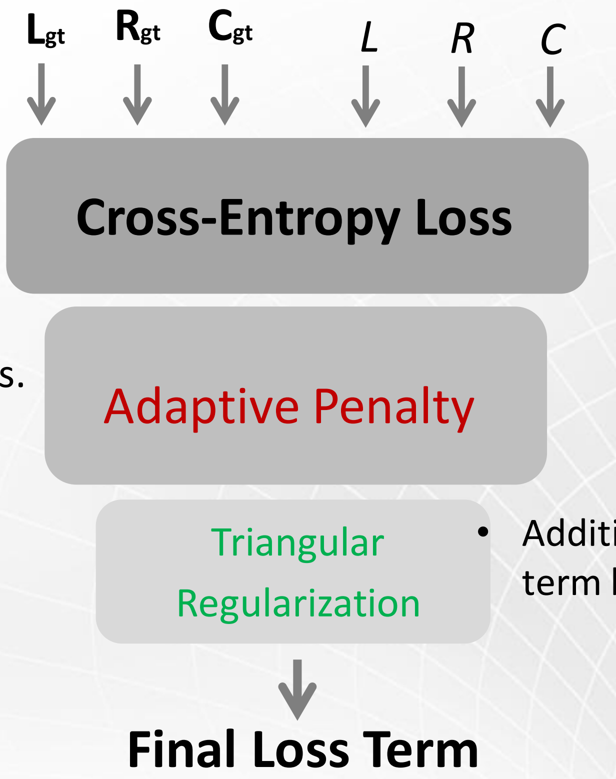


-Information from previous n frames are used as additional input → Valuable time-related priors

Lane ID Estimation – Loss function

Classification task N = 8 with a proposed cost function is composed of 3 parts:

- – Cross Entropy Loss
- – Adaptive Penalty
- – Triangular Regularization



$$L_{cross-entropy}(x, y) = -\sum_i y_i \log(x_i)$$

- More Weights to smaller estimates.

$$1 + \exp(-5 * L)$$
$$1 + \exp(-5 * R)$$

- Additional penalty term based on

R + L + 1 = C

- **L_gt: Ground Truth LEFT Lane ID**
- **R_gt: Ground Truth RIGHT lane ID**
- **C_gt: Ground Truth LANE count**

- *L: Estimated LEFT Lane ID*
- *R: Estimated RIGHT lane ID*
- *C: Estimated LANE count*

Lane ID Estimation – Brightness Pre-processing

- **Adaptive perceived brightness adjustment (Pre-processing)**

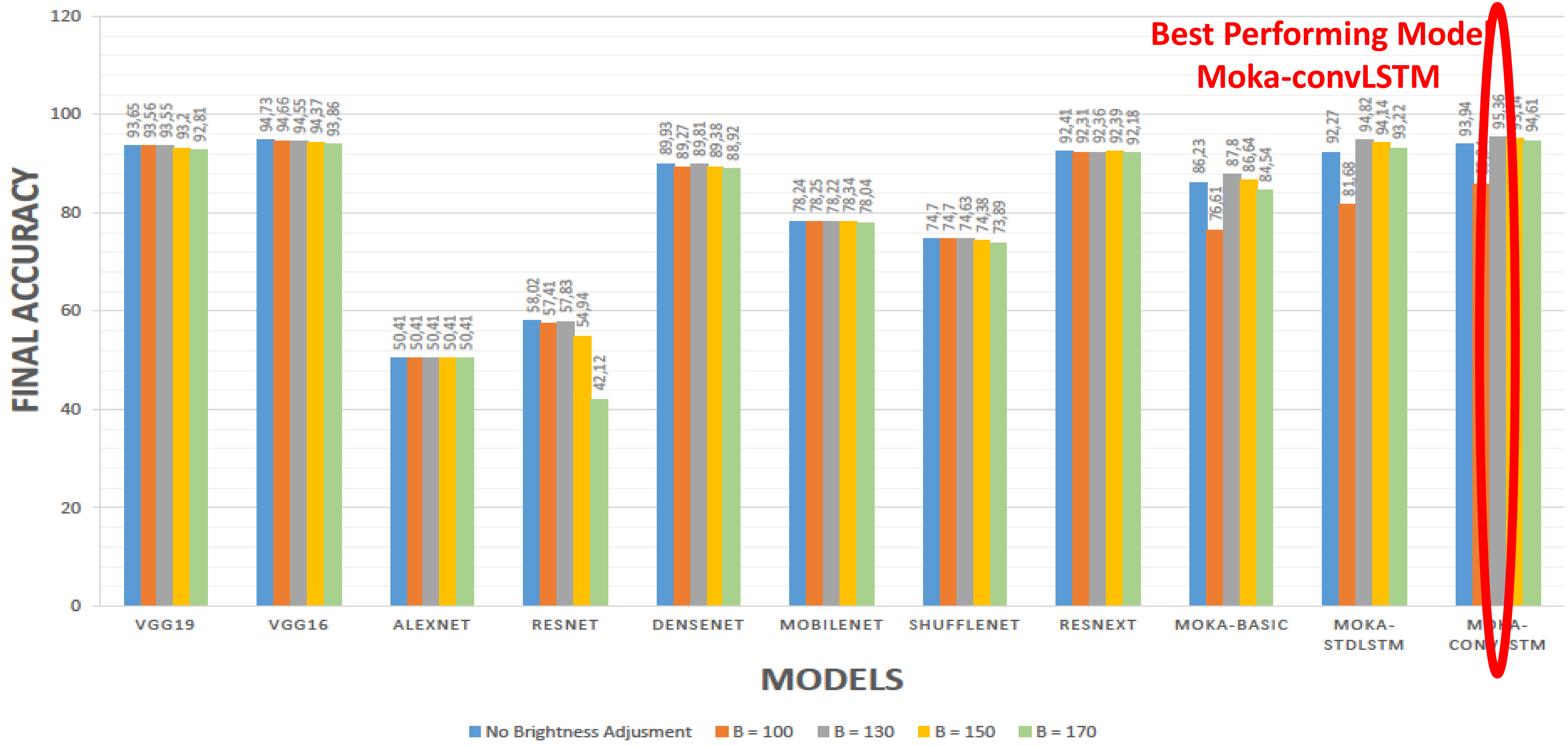
  - we track the average perceived brightness of the driving sequence under consideration.
  - If the perceived brightness of the current frame is below the tracked average (according to a specific threshold) ➔ Adjust the brightness of the frame.
  - The brightness adjustment can be done via a **linear transformation of the pixel intensities**, or using gamma correction (with a corresponding alpha parameter).
  - Optionally, we implemented a new layer in our neural network, which aims at learning the optimal alpha used for the gamma correction

$$R'(x,y) = \min(255, \alpha \,.\, R(x,y))$$

$$G'(x,y) = \min(255, \alpha \,.\, G(x,y))$$

$$B'(x,y) = \min(255, \alpha \,.\, B(x,y))$$

Lane ID Estimation – Brightness Pre-processing



Lane ID Estimation – Decision Module

- The trained model outputs two lane ID candidates according to each convention.
- We need to decide on which output (left or right ID) will be considered as the final estimate.

Model → **L: Left Estimate** / **R: Right Estimate** → **Which estimate to pick as the correct one?**

Left Output:

0.8	**1.4**	0.9	0.7	0.6	0.4	0.4	0.2

Right Output:

0.3	0.7	0.4	**0.9**	0.7	0.3	0.3	0.2

Left or right Lane ID ?

Lane ID Estimation – Best Decision Module

1. Using the classification vectors for each lane ID estimate → select the maximum activation value from left estimate output and right estimate output.
2. From each left and right selected activation value → Subtract the mean activation value of the corresponding vector.
3. The decision about the final output lane ID will be based on the comparison of these 2 values

Left Output Vector:

Lane ID:	1	2	3	4	5	6	7	8
	0.8	**1.4**	0.9	0.7	0.6	0.4	0.4	0.2

Left Max: 1.4 || Left Mean: 0.675

↓

New Left Max: 1.4-0.675 = 0.725

Right Output Vector:

Lane ID:	1	2	3	4	5	6	7	8
	0.3	0.7	0.4	0.9	0.7	0.3	0.3	0.2

Right Max: 0.9|| Right Mean: 0.475

↓

New Right Max: 0.9-0.475 = 0.425

↓

0.725 > 0.425 → Left ID (2) is selected

Lane ID Estimation – Decision Module

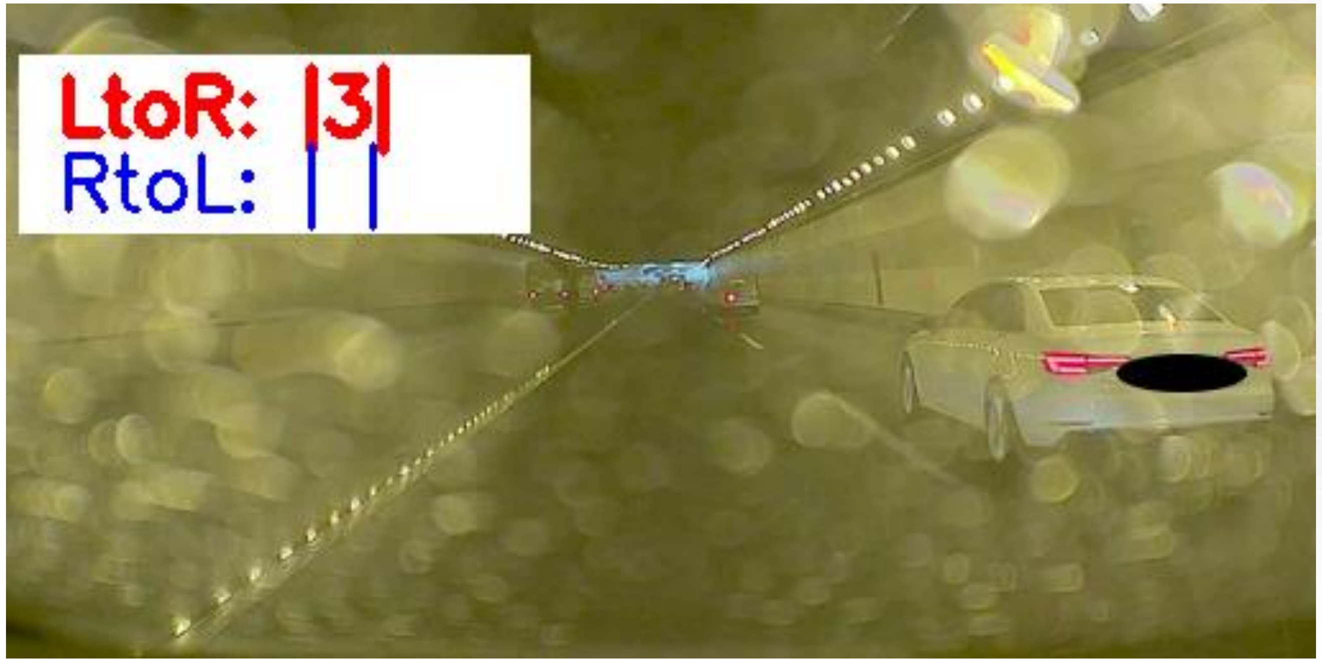
- We apply different decision modules with different architectures to compare final lane ID estimation performance.

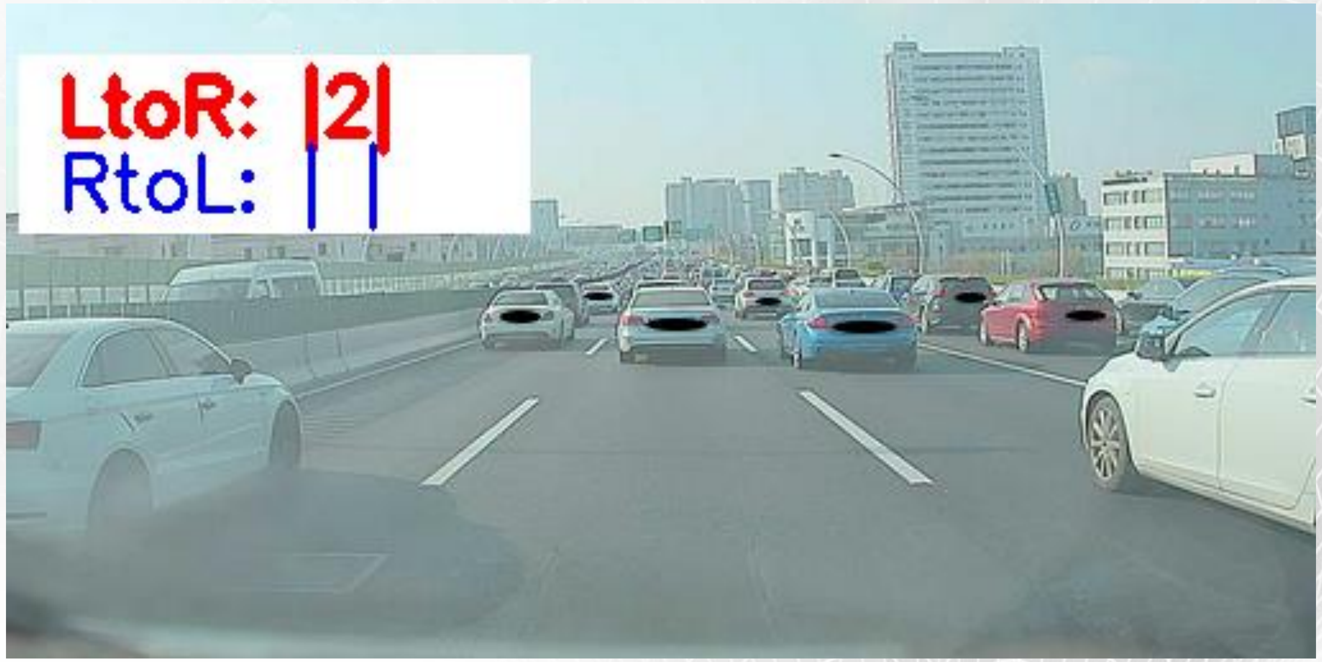
The decision corresponds to the final choice we take to pick up the best convention to use between left and right.

Model	Max	Max-M	E	Max-E	Z-score
VGG19	93.43 %	93.65 %	81.86 %	93.37 %	94.22 %
VGG16	94.47 %	94.73 %	83.23 %	92.27 %	94.95 %
Alexnet	49.84 %	50.41 %	47.02 %	49.14 %	49.83 %
Resnet	40.67 %	58.02 %	30.29 %	49.89 %	43.93 %
Densenet	89.99 %	89.93 %	73.37 %	89.11 %	90.24 %
Mobilenet	78.06 %	78.24 %	57.96 %	75.24 %	77.44 %
Shufflenet	74.35 %	74.70 %	53.80 %	74.68 %	74.24 %
ResNext	92.39 %	92.41 %	81.76 %	88.61 %	91.79 %
MOKA-basic	86.11 %	86.23 %	67.37 %	70.56 %	82.35 %
MOKA-StdLSTM-B130	92.21 %	94.82 %	76.04 %	89.88 %	92.33 %
MOKA-convLSTM-B130	95.47 %	95.36 %	83.09 %	84.21 %	95.02 %

PERFORMANCE COMPARISON (FINAL ACCURACY) USING DIFFERENT DECISION CRITERIA FOR THE CHOICE BETWEEN LEFT AND RIGHT CONVENTIONS.

Lane ID Estimation – Sample Results from our dataset





Lane ID Estimation – Sample Results on random videos

Lane ID Estimation – Sample Results on random videos

Lane ID Estimation – Sample Results on random videos

Lane ID Estimation – Conclusion

- We perform lane ID estimation for autonomous driving using CNNs (localization, mapping…).
- We propose a real-time vision only based solution (monocular) to predict lane ID.
- The solution is targeting low-complexity and limited runtime requirements for real-world autonomous driving scenarios.
- We harnesses the temporal dimension inherent to the input sequences to improve upon high complexity state-of-the-art models.
- We achieve more than 95% accuracy on a challenging test set with extreme conditions and different routes.
- We visually verify the performance of our lane ID model with random videos downloaded from the internet.

Thank You

Ibrahim.Halfaoui@huawei.com