Automated Harvesting and Convolutional Classification of Road-Marking Training Data from HD-Map Vectors

Training Dataset One · OpenFeatureX harvester + FeatureCNN

AG2I — Boulder, Colorado · 9 June 2026

Abstract

A road-marking training set was generated without manual annotation by projecting HD-map road-marking vector layers onto a co-registered nadir true-orthophoto (0.076 m ground sample distance) and cropping one image chip per feature, labelled by the feature's coded class attribute mapped to the OpenFeatureX catalogue. From 13 633 polygon and 47 483 line features, 15 000 chips were retained under a 1 000-per-class cap across 26 classes. A four-block convolutional classifier attains top-1 accuracy 58.6 %, mean per-class recall 62.6 %, and mean per-class precision 53.1 % on a 2 993-chip held-out split. In a controlled ablation — identical resolution, data, and schedule, varying only the network head — replacing global average pooling with a 4 × 4 spatial head raises top-1 from 49.7 % to 58.6 % and right-turn-arrow recall from 28 % to 65 % (left arrow 45 % to 70 %), isolating chirality, the mirror relationship between left and right arrows that pooling cannot resolve, as the dominant prior error; 18 of 25 classes improve. On contiguous road scenes, where high-frequency classes dominate, classification accuracy is 56–95 %. Labelling effort was zero.

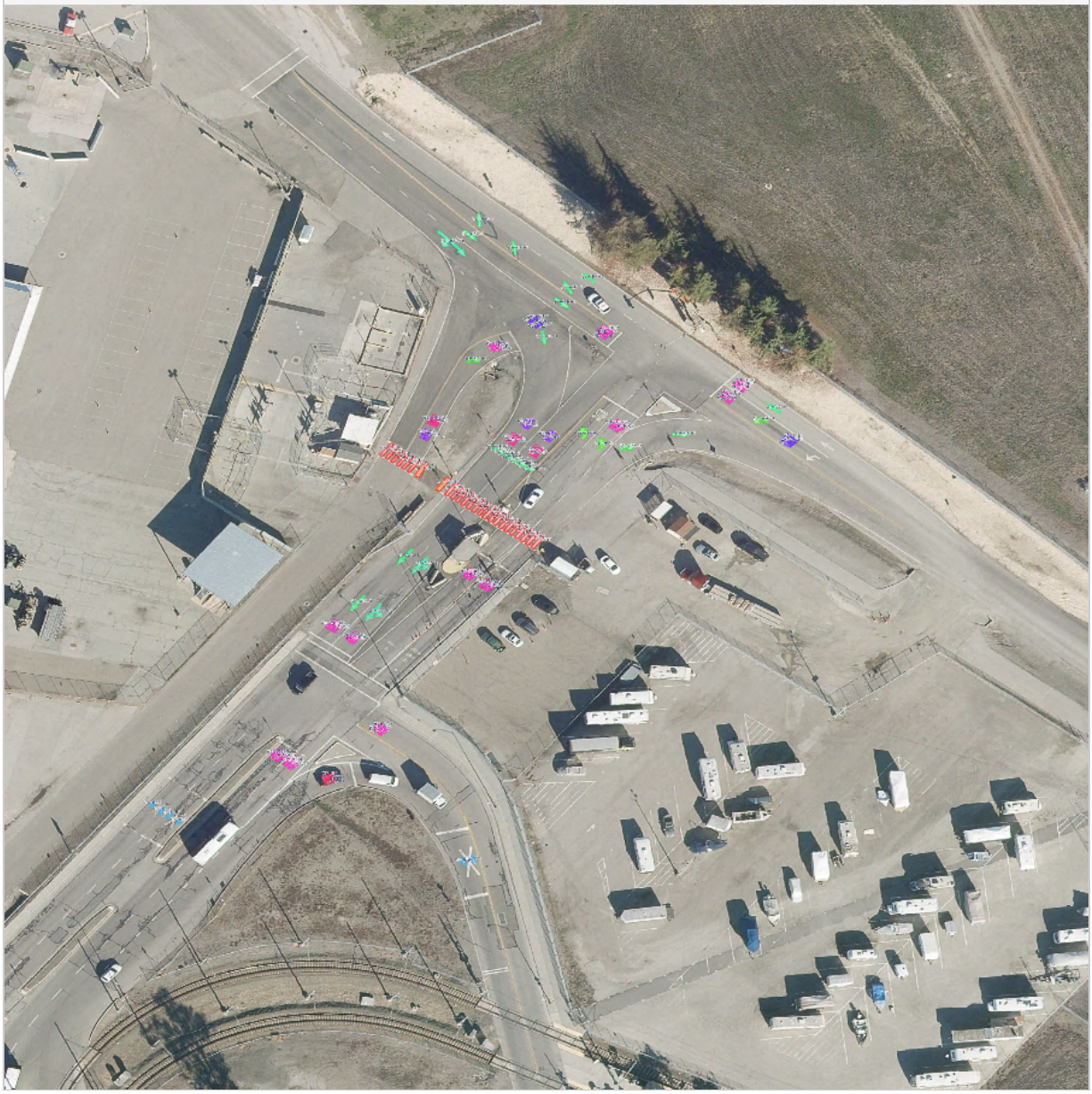
1. Data

- Imagery: nadir true-orthophoto, 0.076 m ground sample distance, three-band, geographic coordinate reference, 354 150 × 185 881 px.
- Labels: two co-registered HD-map vector layers — a road-marking polygon layer (13 633 features: arrows, word markings, crosswalks) and a road-marking line layer (47 483 features: longitudinal line types). `Function` and `Class` stored as integer codes; decoded through the source coded-value domains.
- Annotation: none. The HD-map vectors are the labels.

2. Method

Harvesting. Integer `Function/Class` codes were decoded through the source domains and mapped to OpenFeatureX catalogue codes. Each feature was reprojected to the orthophoto coordinate reference. Polygon features were cropped at their envelope (+30 %); line features in a fixed 4 m window centred on a mid-line vertex. Chips were resized to 96 px. A 1 000-per-class cap bounded over-represented classes. Retained chips: 15 000 across 26 classes.

Figure B. HD-map marking vectors projected onto the orthophoto (146 labeled features)



Network. A four-block convolutional network (32–64–128–128 channels, batch normalisation, max pooling). The head replaces global average pooling with a 4 × 4 adaptive-average spatial head before the linear classifier. Global average pooling is invariant to translation and orientation but discards spatial arrangement; arrangement is the discriminator for chirality (left versus right arrows), word letter layout, and line dash patterns, so a coarse spatial map is retained.

Training. Stratified 80/20 split (train 12 007, validation 2 993), held fixed. Augmentation: cardinal and ±45° rotation (no mirroring, which inverts arrow chirality), scale 0.82–1.18, brightness/contrast jitter,

30 % Gaussian blur. Class-balanced sampling; Adam at 1.5×10^{-3} with cosine annealing to zero; weight decay 1×10^{-4}; label smoothing 0.1; 30 epochs; cross-entropy loss.

3. Results

Improvement from the spatial head (controlled). Holding chip resolution, data, and training schedule fixed and changing only the network head, the 4 × 4 spatial head raises top-1 accuracy from 49.7 % to 58.6 %, mean per-class recall from 53.6 % to 62.6 %, mean per-class precision from 42.1 % to 53.1 %, and — the largest single prior error — right-turn-arrow recall from 28 % to 65 %.

Controlled ablation: a global-average-pooling head versus a 4 × 4 spatial head, both at 96 px on the fixed held-out split with all other settings identical, so each difference is attributable to the head alone.

Head	Top-1	Macro recall	Macro precision	Right-arrow recall	Left-arrow recall
Global average pool	49.7 %	53.6 %	42.1 %	28 %	45 %
Spatial 4 × 4	**58.6 %**	**62.6 %**	**53.1 %**	**65 %**	**70 %**

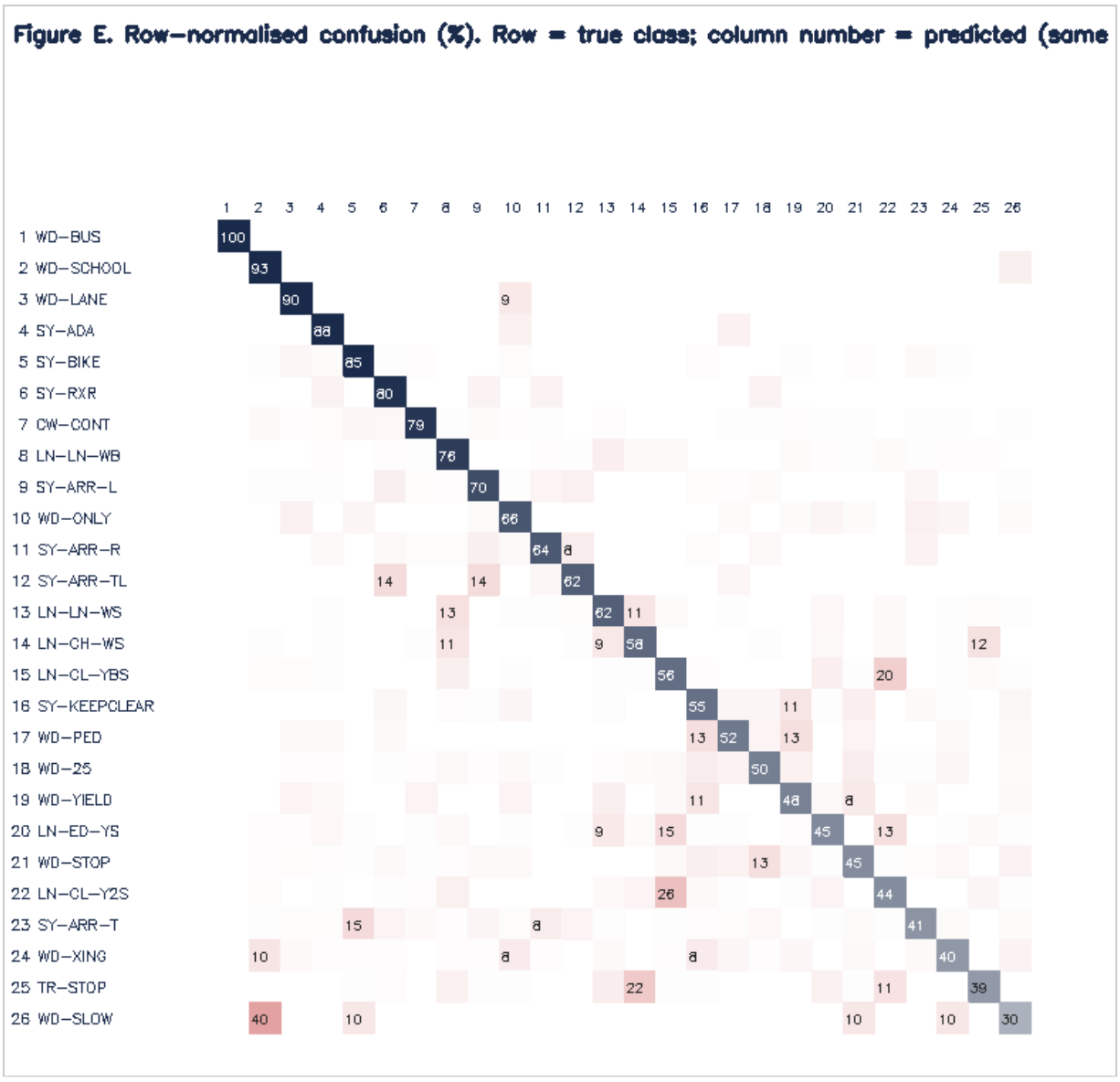
Global average pooling collapses the final feature map to a single vector per channel, discarding where each feature lies; the spatial head retains a 4 × 4 grid. The effect is largest exactly where arrangement is the discriminator: the right turn arrow gains 37 points (28 %→65 %) and the left arrow 25 points (45 %→70 %), the two being mirror images that pooling cannot separate. The gain is broad — 18 of 25 classes improve — and lifts top-1 by 8.9 points.

Figure F. Classifier on real markings: 60/103 = 58% correct (green) vs mismatch (red)

Per-class validation recall and precision (spatial-head model), with the change in recall versus the global-pool control (Δ):

OFX code	Class	Recall	Precision	Δ vs global pool	n
WD-SCHOOL	SCHOOL (word)	93 %	30 %	−7	15
WD-LANE	LANE (word)	91 %	24 %	+18	11
SY-ADA	Accessibility symbol	88 %	31 %	+0	17
SY-BIKE	Bicycle symbol	86 %	79 %	+2	200

OFX code	Class	Recall	Precision	Δ vs global pool	n
SY-RXR	Railroad crossing	81 %	25 %	+14	21
CW-CONT	Crosswalk	79 %	88 %	+11	200
LN-LN-WB	Broken white lane line	76 %	60 %	+0	200
SY-ARR-L	Left turn arrow	70 %	81 %	**+25**	200
WD-ONLY	ONLY (word)	66 %	45 %	+20	50
SY-ARR-R	Right turn arrow	65 %	59 %	**+36**	74
SY-ARR-TL	Through-left arrow	63 %	34 %	+22	27
LN-LN-WS	Solid white lane line	62 %	62 %	−2	200
LN-CH-WS	Channelizing line	58 %	60 %	+16	185
LN-CL-YBS	Broken yellow centre line	56 %	52 %	+3	200
SY-KEEPCLEAR	Keep clear	56 %	53 %	+11	124
WD-PED	PED XING (word)	52 %	37 %	+9	44
WD-25	Speed legend	50 %	63 %	+15	168
WD-YIELD	YIELD (word)	49 %	30 %	+2	45
LN-ED-YS	Yellow solid edge line	46 %	65 %	−0	200
WD-STOP	STOP (word)	45 %	68 %	+4	200
LN-CL-Y2S	Double yellow centre line	44 %	48 %	+4	200
SY-ARR-T	Through arrow	42 %	73 %	+12	200
WD-XING	XING (word)	40 %	53 %	+11	82
TR-STOP	Stop line	39 %	51 %	+14	119
WD-SLOW	SLOW (word)	30 %	8 %	−10	10



4. Confusion structure

With chirality addressed, the residual error (Figure E) is dominated by three mechanisms. (i) Shape similarity: the through arrow (42 %) is still taken for the bicycle symbol (31 of 200), both elongated forms. (ii) Short-word ambiguity: speed legend ↔ STOP (27), XING ↔ SCHOOL, SLOW ↔ SCHOOL — short white words on dark asphalt at 0.076 m. (iii) Line-type pairs: double ↔ broken yellow (53, 41) and broken ↔ solid white (26, 14), which differ only in dash pattern. Precision remains lowest where a high-recall minority class over-predicts onto a frequent class (SY-RXR, WD-SCHOOL, WD-SLOW), reflected in the macro precision (53.1 %). The right and left turn arrows, the prior dominant confusion, are now separated (65 % and 70 % recall).

5. Conclusion

Decoding and projecting two HD-map road-marking vector layers onto a 0.076 m orthophoto produced 15 000 labelled chips across 26 classes with no manual annotation. In a controlled ablation varying only the network head, replacing global average pooling with a 4 × 4 spatial head raised top-1 accuracy from 49.7 % to 58.6 % and right-turn-arrow recall from 28 % to 65 % — resolving the chirality error that global pooling caused, with 18 of 25 classes improving. On whole road scenes the classifier reaches 56–95 %. The residual error is structured (shape-similar forms, short-word ambiguity, dash-pattern line pairs), which localises the next improvements. The procedure is annotation-free and scales with HD-map coverage.

References

1. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86(11), 2278–2324.
2. Lin, M., Chen, Q., Yan, S. (2014). Network in network. *ICLR*.
3. Ioffe, S., Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *ICML*, 448–456.
4. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z. (2016). Rethinking the Inception architecture for computer vision. *CVPR*, 2818–2826.
5. Loshchilov, I., Hutter, F. (2017). SGDR: stochastic gradient descent with warm restarts. *ICLR*.
6. Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. B* 36(2), 111–147.