

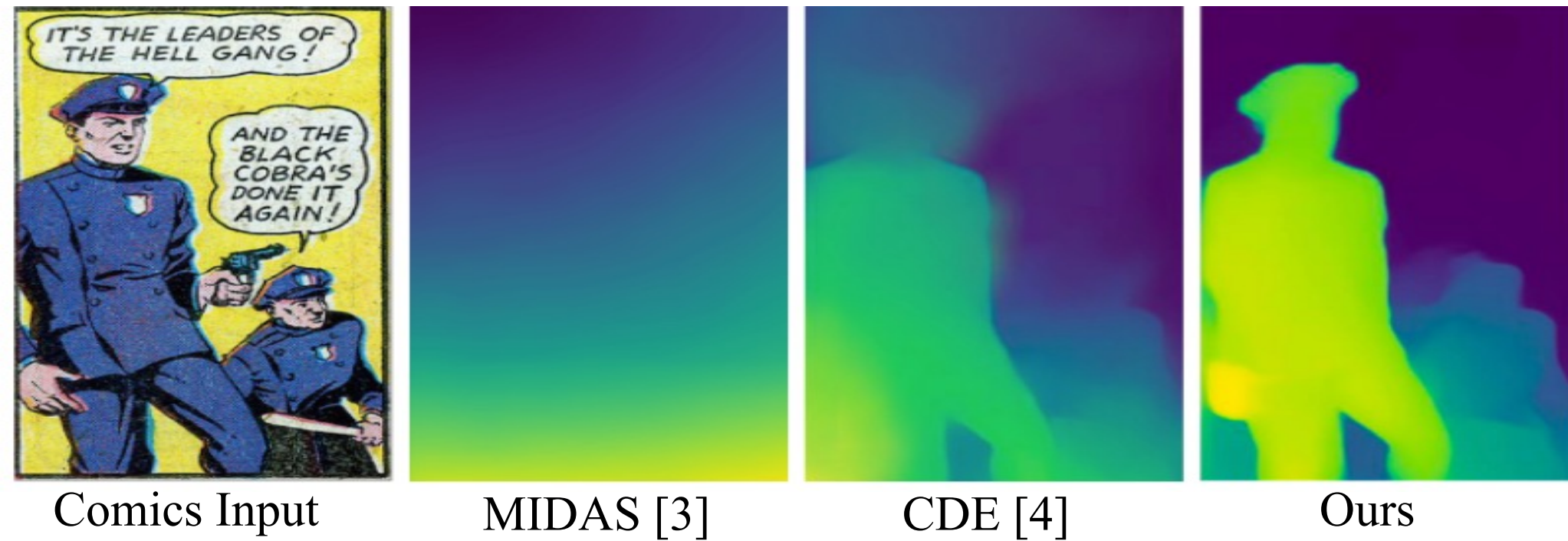


## Motivation

Estimating depth in the comics domain is subject to many challenges, including a) occlusions between characters; b) unusual object sizes (the bird here); c) unusual perspective; d) and e) different illustrative styles. The images are monocular, lack ground-truth depth annotations and are noisy. We thus, use an off-the-shelf unsupervised image to image translation method to translate the comics images to natural ones and then use an attention-guided monocular depth estimator to predict their depth.



Challenges in the comics domain.



The state-of-the-art monocular depth estimation models fail to predict accurate depth when directly employed on comics images.

## Contributions

We introduce a cross-domain depth estimation method by leveraging an off-the-shelf unsupervised I2I translation method.

We exploit the contextual information for depth prediction of a given scene where we use an inner feature-based GAN and a Laplacian edge detector.

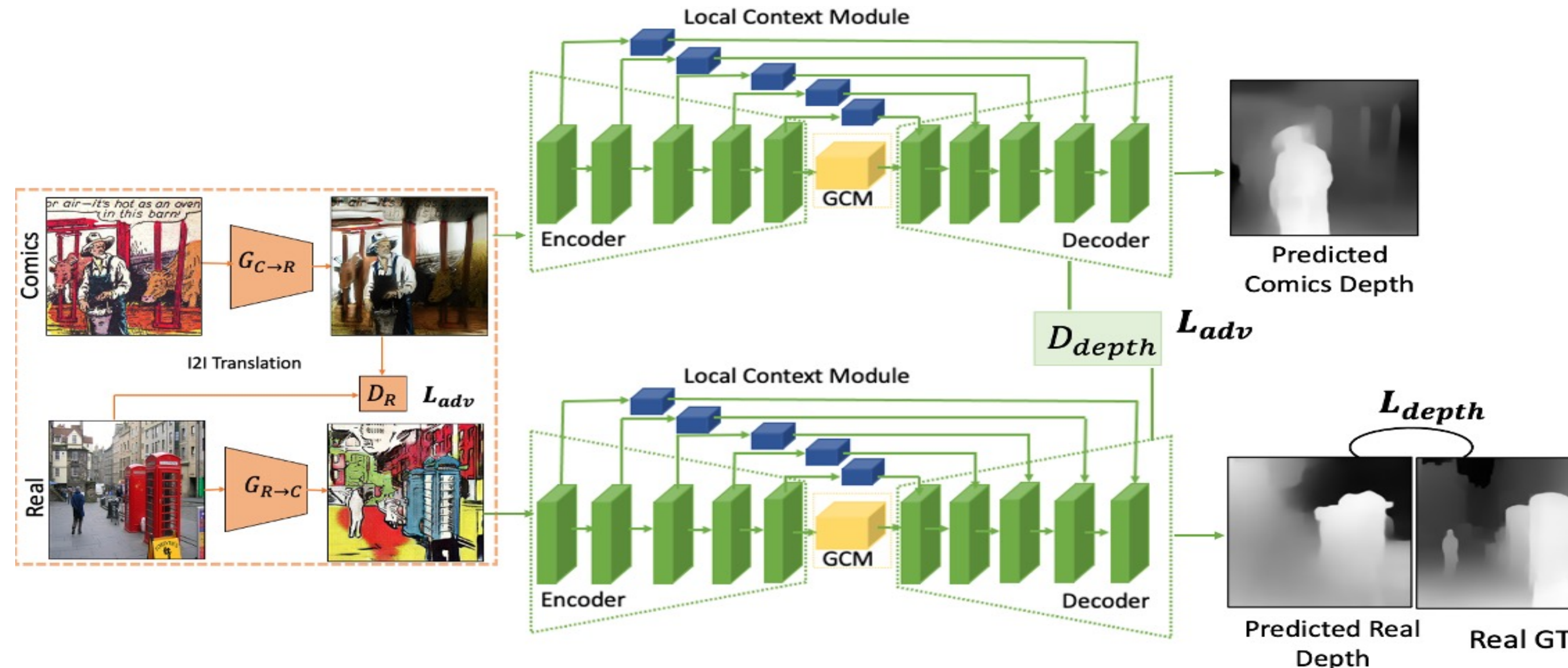
By introducing a text detector, we reduce the artefacts from text and speech balloons in the depth predictions, which are specific to comics.

We introduce a benchmark dataset for comics images with 450 manually annotated image-depth pairs.

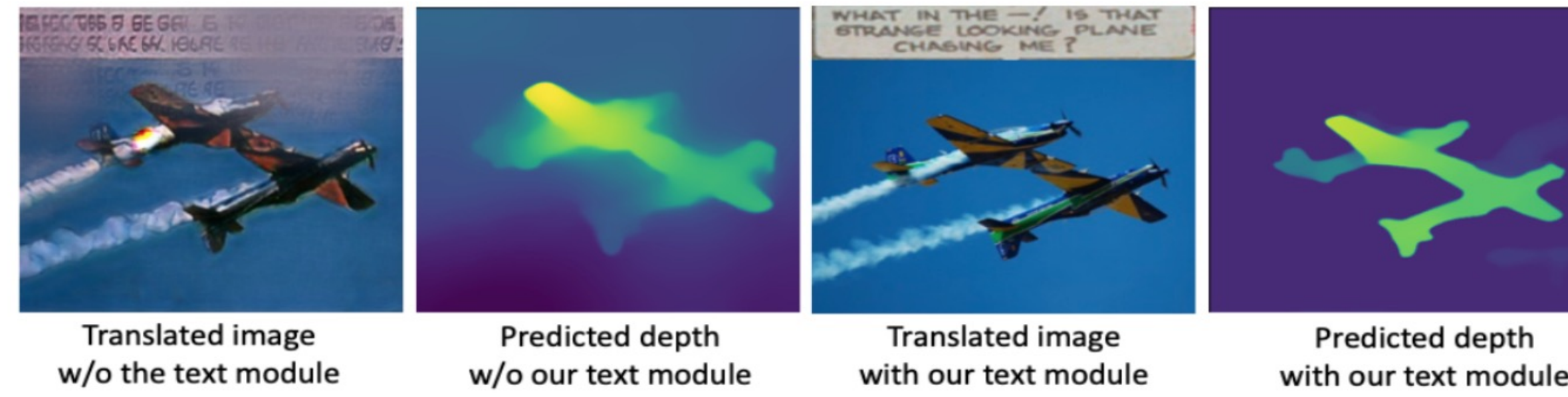


Example from our annotated benchmark for comics images.

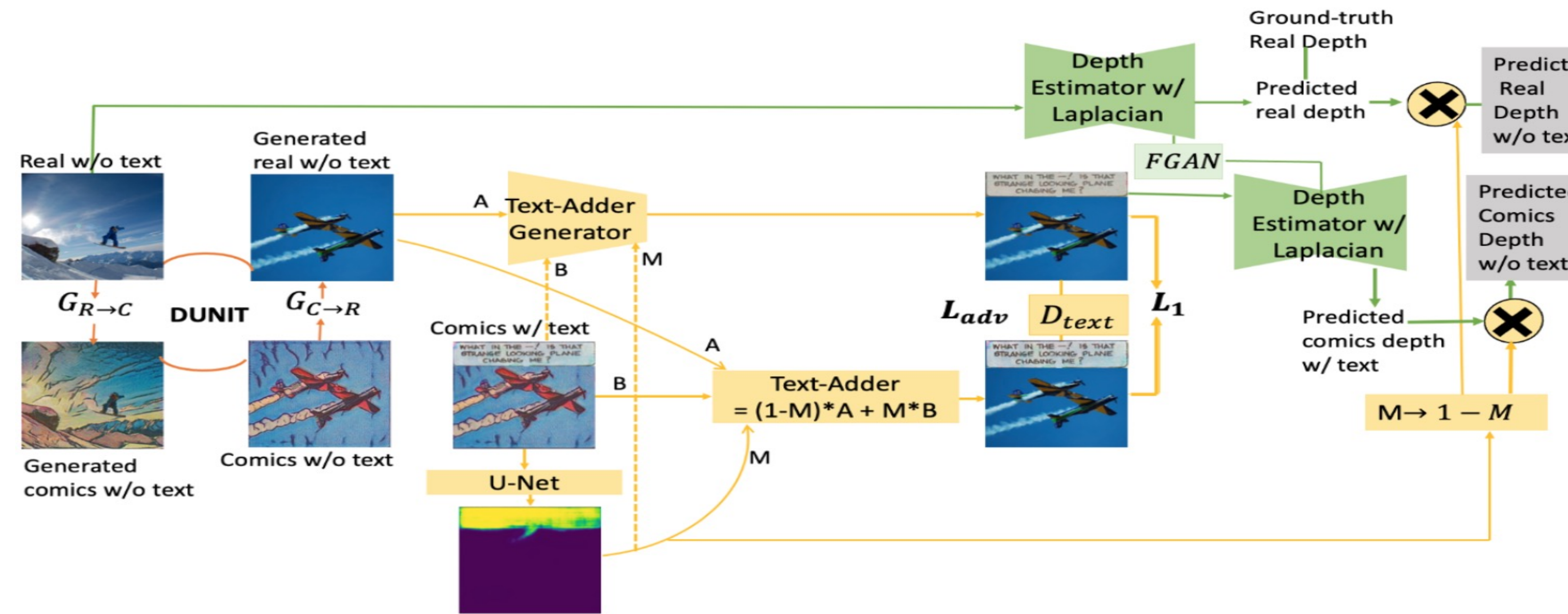
## Method Overview



## Text Detection Module

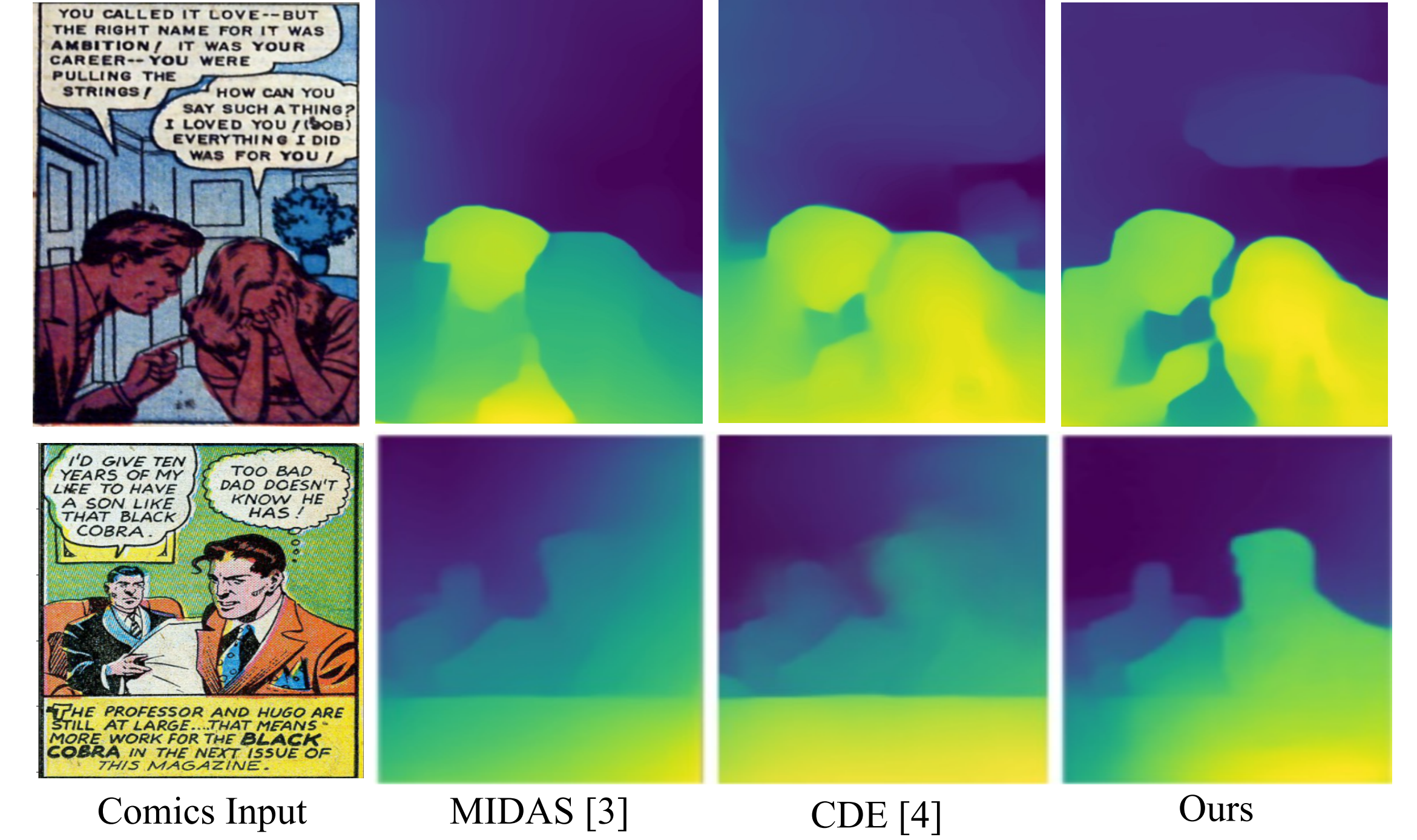


Motivation for our text module.



Overview of our approach with the text module.

## Depth Results for Comics Images



## Quantitative Comparison

Method	DCM Validation Images				eBDtheque Validation Images			
	AbsRel↓	SqRel↓	RMSE↓	RMSE log↓	AbsRel↓	SqRel↓	RMSE↓	RMSE log↓
T2Net [1]	0.351	0.416	1.117	0.415	0.491	0.555	1.459	0.777
Song et.al. [2]	0.339	0.401	1.098	0.402	0.479	0.520	1.431	0.711
MIDAS [3]	0.309	0.381	1.033	0.375	<u>0.419</u>	<u>0.503</u>	1.416	0.659
CDE [4]	<u>0.304</u>	<u>0.374</u>	<u>1.024</u>	<u>0.367</u>	0.424	0.511	<u>1.415</u>	<u>0.647</u>
Ours	<b>0.251</b>	<b>0.318</b>	<b>0.971</b>	<b>0.305</b>	<b>0.376</b>	<b>0.448</b>	<b>1.364</b>	<b>0.553</b>

## References

- [1] T2net: Synthetic-to-realistic translation for solving single-image depth estimation tasks, ECCV, 2018.
- [2] Monocular depth estimation using laplacian pyramid-based depth residuals, IEEE Transactions on Circuits and Systems for Video Technology, 2021.
- [3] Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020.
- [4] Leveraging contextual information for monocular depth estimation, IEEE Access, 8, 2020.
- [5] Digital comics image indexing based on deep learning Journal of Imaging, 4(7), 2018.
- [6] ebdtheque: a representative database of comics, ICDAR, 2013.

**Acknowledgement:** This work was supported in part by the Swiss National Science Foundation via the Sinergia grant CRSII5-180359.