CorNetv2: Deep Learning for Mosaicking Maize Fields

Dewi Endah Kharismawati, Filiz Bunyak, Rumana Aktar, Kannappan Palaniappan, Toni Kazic

University of Missouri

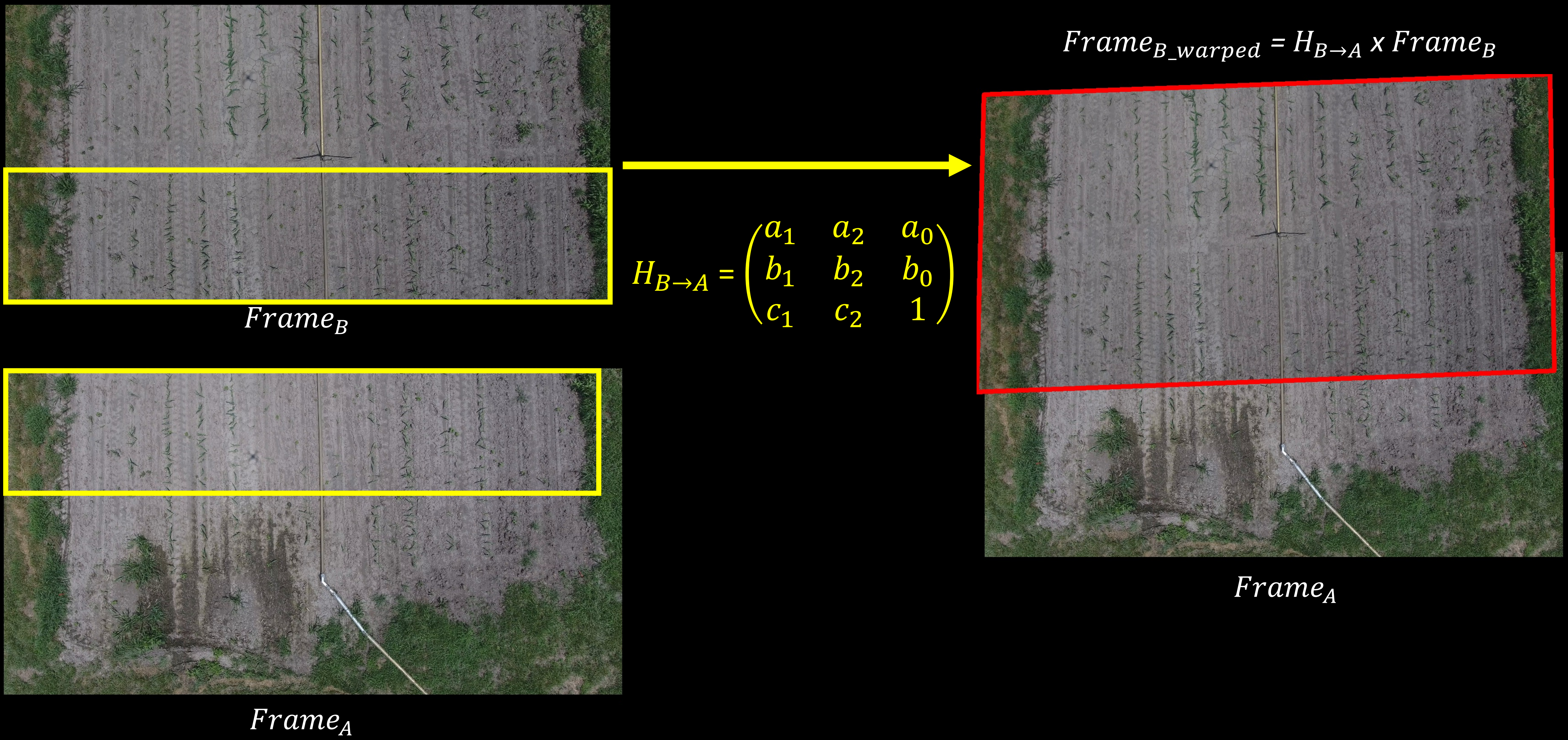
Motivation

- Working in the field is often challenging and labor intensive
- Small, low-cost drones could provide automated, non-destructive phenotyping and stress monitoring
- Dynamic scouting for phenotypes would be more efficient and less laborious, but requires new approaches to mosaicking and flying
- Current flight trajectories take technical resources not widely available to researchers and farmers, and large fields need multiple flights
- Mosaicking aerial imagery is the first step in constructing a high-resolution view of a field

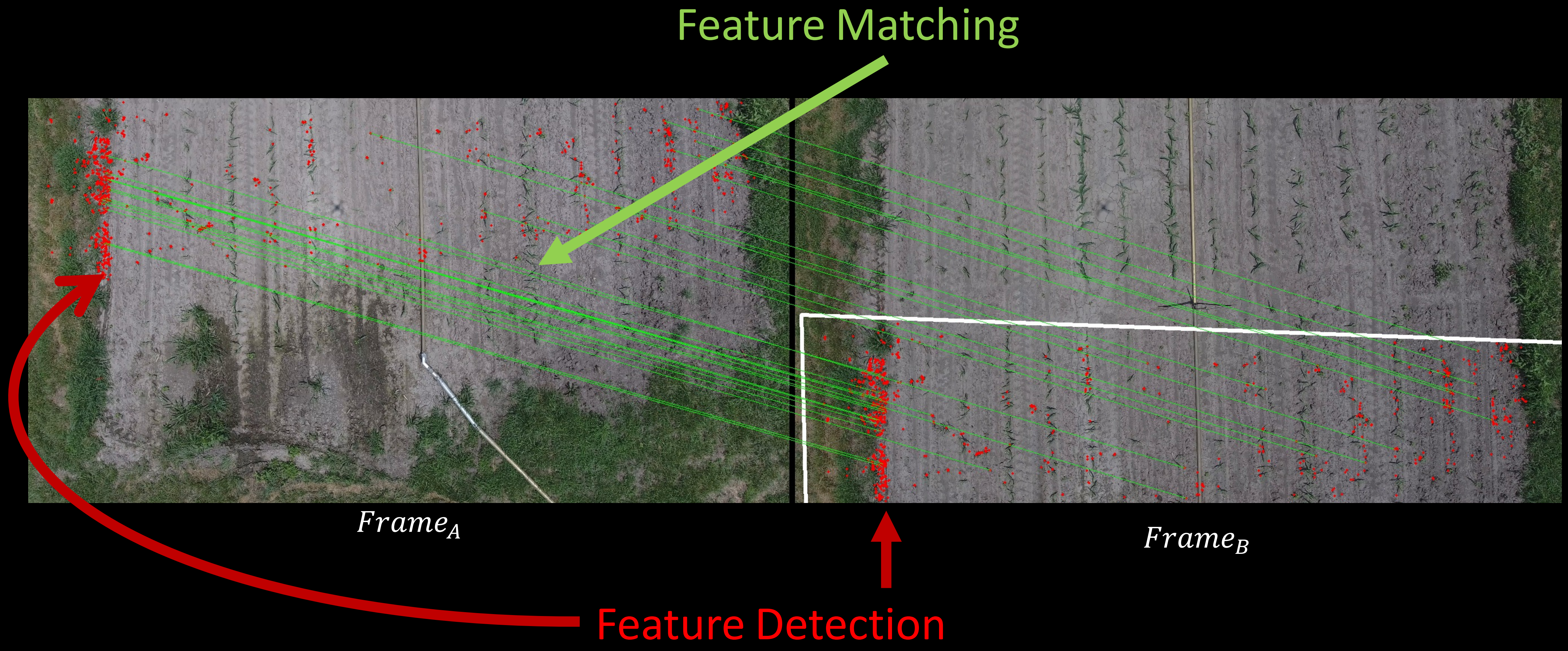
Mosaicking Challenges and Our Solution

- Agricultural imagery is much more challenging because of the repetitive and irregular features
- Many current methods are computationally expensive and rely on metadata from expensive drones, preplanned flights, and ground control points for registration
- We are building *CorNetv2, our unsupervised deep learning model* to mosaic maize fields
  - Efficient and accurate homography estimation without metadata
  - Uses images from cheap drones
  - No preplanned flights and handles different camera poses for richer phenotyping

An Overview of Mosaicking



Traditional Mosaicking Only Looks at Points

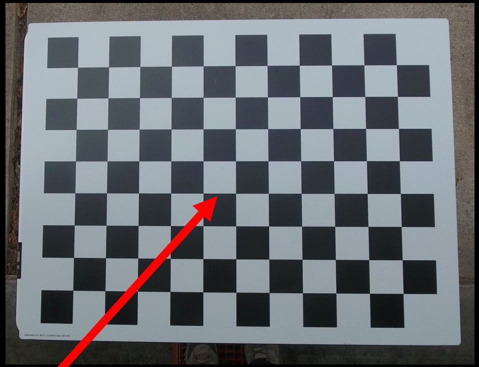


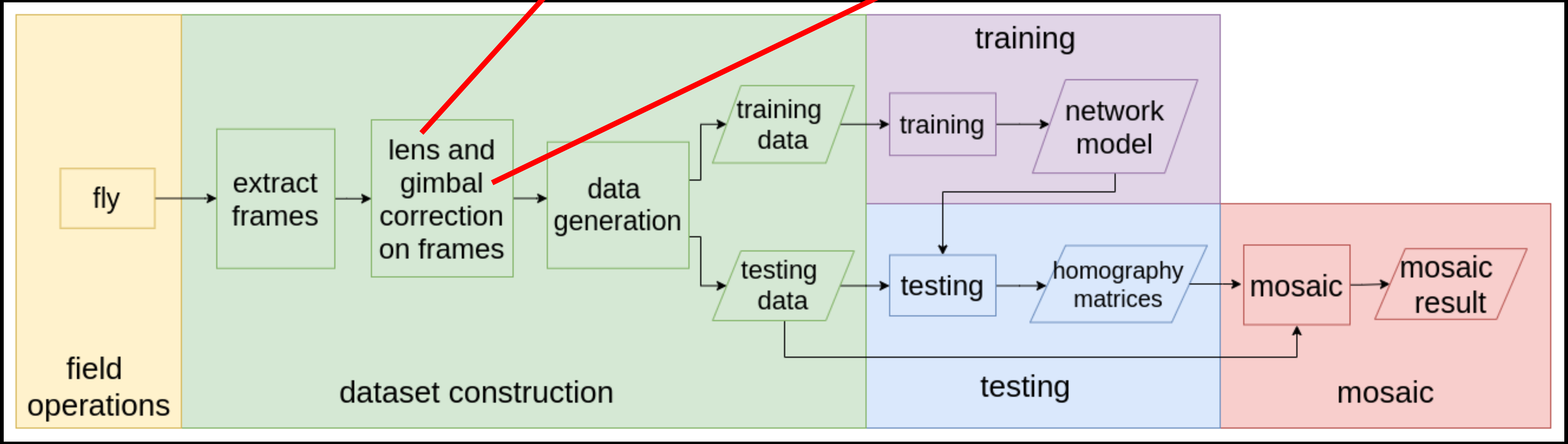
Poor regression -- confusion in matching -- produces erroneous mosaics.

5 Regression Models for Homography Estimation

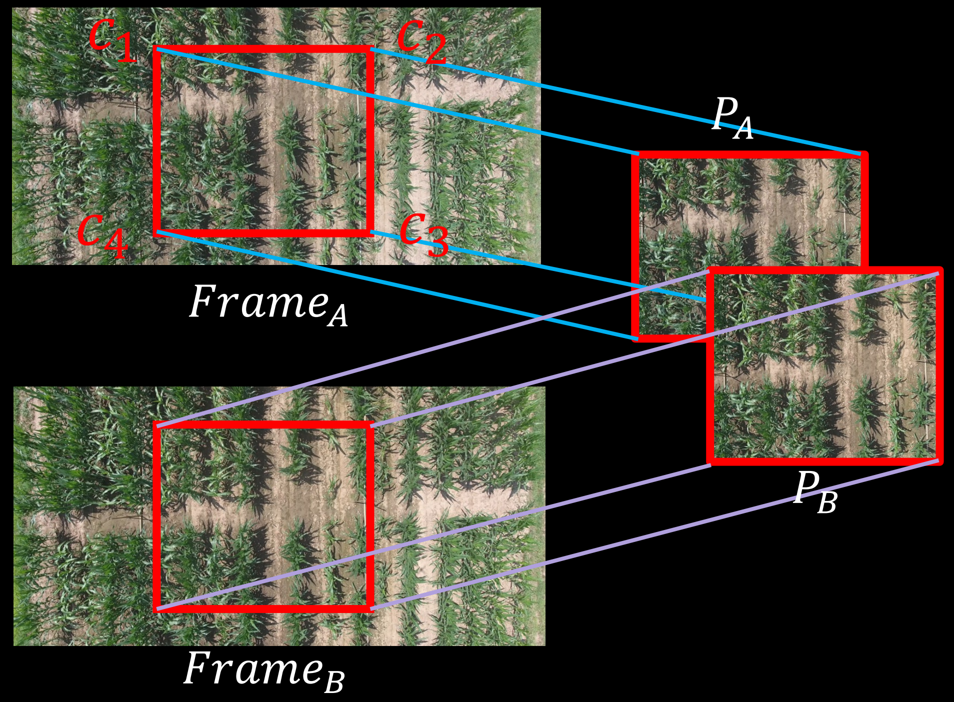
1. ASIFT: regression over pairs of points; brute force feature matching is expensive and slow
2. Unsupervised Deep Homography *(UDH)*: deep learning model regresses on small patches in non-agricultural imagery, input is 128 x 128 pixels
3. Retrained Unsupervised Deep Homography *(rUDH):* their model, our agricultural imagery
4. *CorNetv1:* our model, our agricultural imagery; regressing on bigger patches with some different movements; input is 512 x 512 pixels, providing more feature information
5. *CorNetv2:* our model; our data are now better balanced for drone movements and corn growth stages; input is 512 x 512 pixels

CorNet's Pipeline





CorNet's Homography Estimation

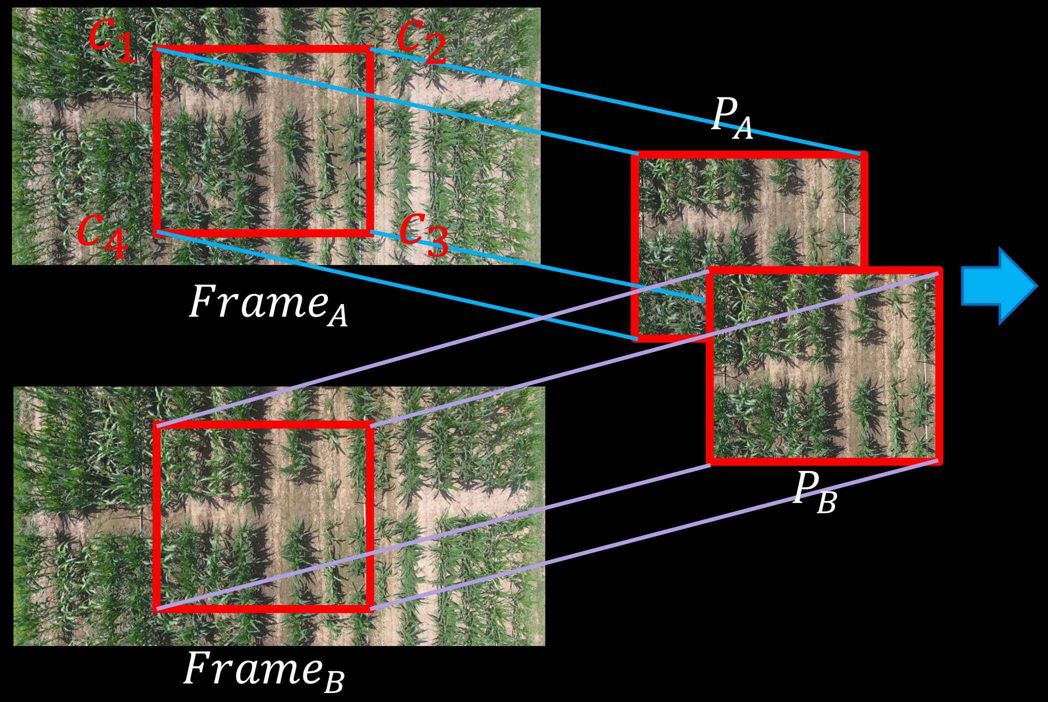


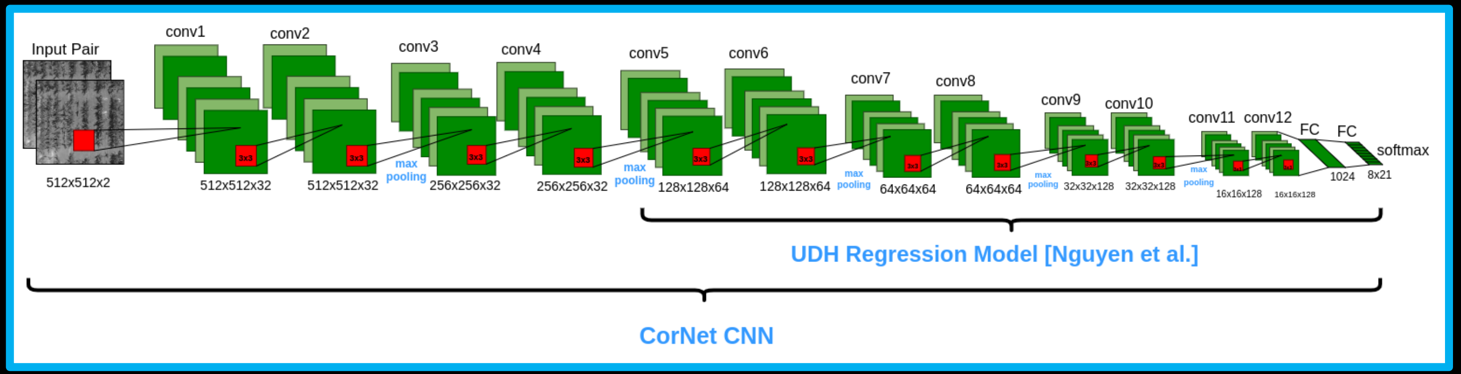
Kharismawati, D. E.; Aliakbarpour, H.; Aktar, R.; Bunyak, F.; Palaniappan, K. & Kazic, T. Bartoli, A. & Fusiello, A. (Eds.) CorNet: unsupervised deep homography estimation for agricultural aerial imagery. *European Conference on Computer Vision 2020 Workshops, part VI: Computer Vision Problems in Plant Phenotyping (CVPPP 2020)*, Springer, 2020, pp. 402-419

CorNet's Homography Estimation

Simple ⟶ Complex

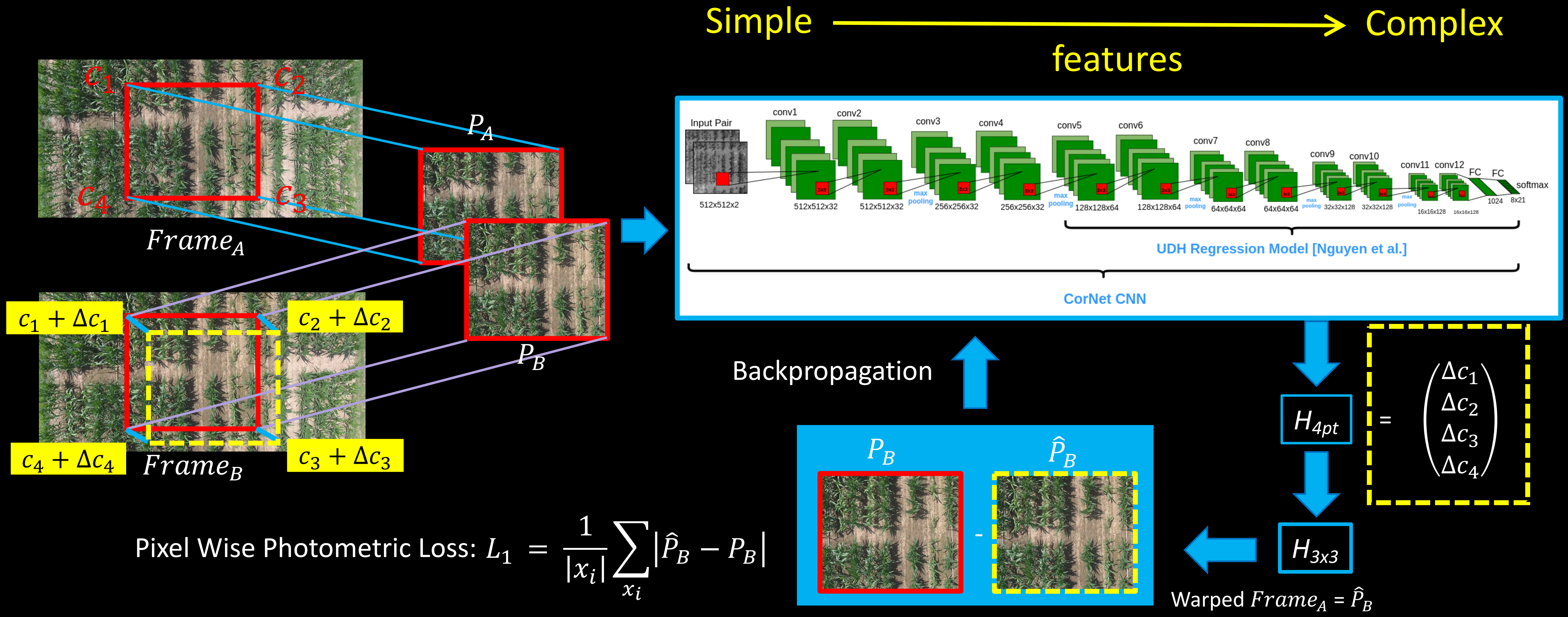
features





Kharismawati, D. E.; Aliakbarpour, H.; Aktar, R.; Bunyak, F.; Palaniappan, K. & Kazic, T. Bartoli, A. & Fusiello, A. (Eds.) CorNet: unsupervised deep homography estimation for agricultural aerial imagery. *European Conference on Computer Vision 2020 Workshops, part VI: Computer Vision Problems in Plant Phenotyping (CVPPP 2020)*, Springer, 2020, pp. 402-419

CorNet's Homography Estimation



Kharismawati, D. E.; Aliakbarpour, H.; Aktar, R.; Bunyak, F.; Palaniappan, K. & Kazic, T. Bartoli, A. & Fusiello, A. (Eds.) CorNet: unsupervised deep homography estimation for agricultural aerial imagery. *European Conference on Computer Vision 2020 Workshops, part VI: Computer Vision Problems in Plant Phenotyping (CVPPP 2020)*, Springer, 2020, pp. 402-419

Training Data: Complex Drone Movements and Corn Growth Stages

Adult: Rotation

Adult: Left slide

Seedling: Clockwise Rotation

Adult: Forward

Adult: Backward

Seedling: Scale

Training datasets

- 27,228 pairs of images for *rUDH*
- 18,422 pairs of images for *CorNetv1 (de novo)*
- 10,800 pairs of images for *CorNetv2* (finetuning v1)

Test Dataset

Evaluate the accuracy on complex movements and generalizability to different landscapes, including landscapes with very few features (farm pond).

Farm pond

Forest Summer

Farm Buildings

UMDC, Urban Buildings

Forest Fall

A Demanding Test: Forward and Backward Trajectories

ASIFT

UDH

rUDH

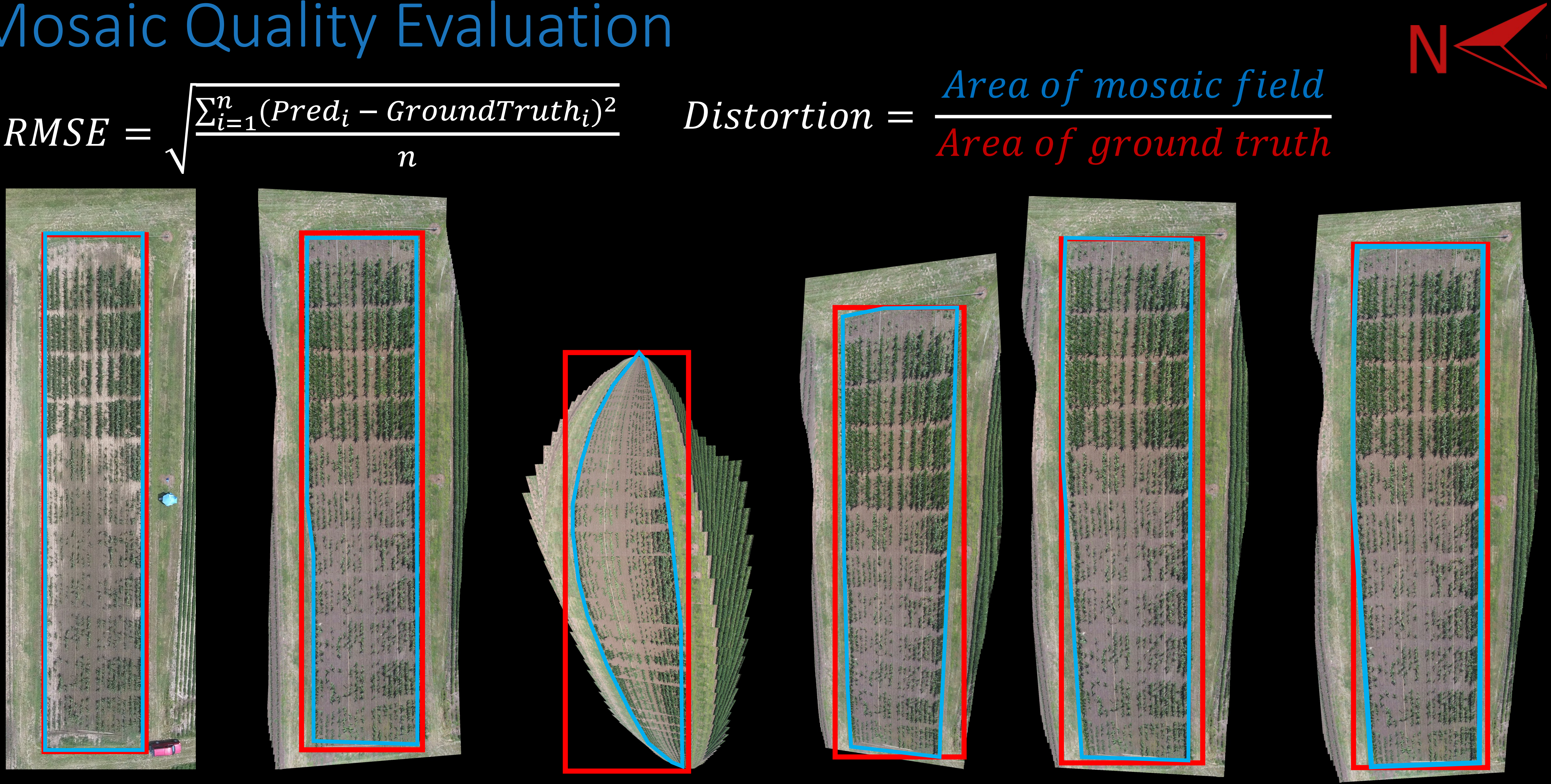
CorNetv1

CorNetv2

Mosaic Quality Evaluation

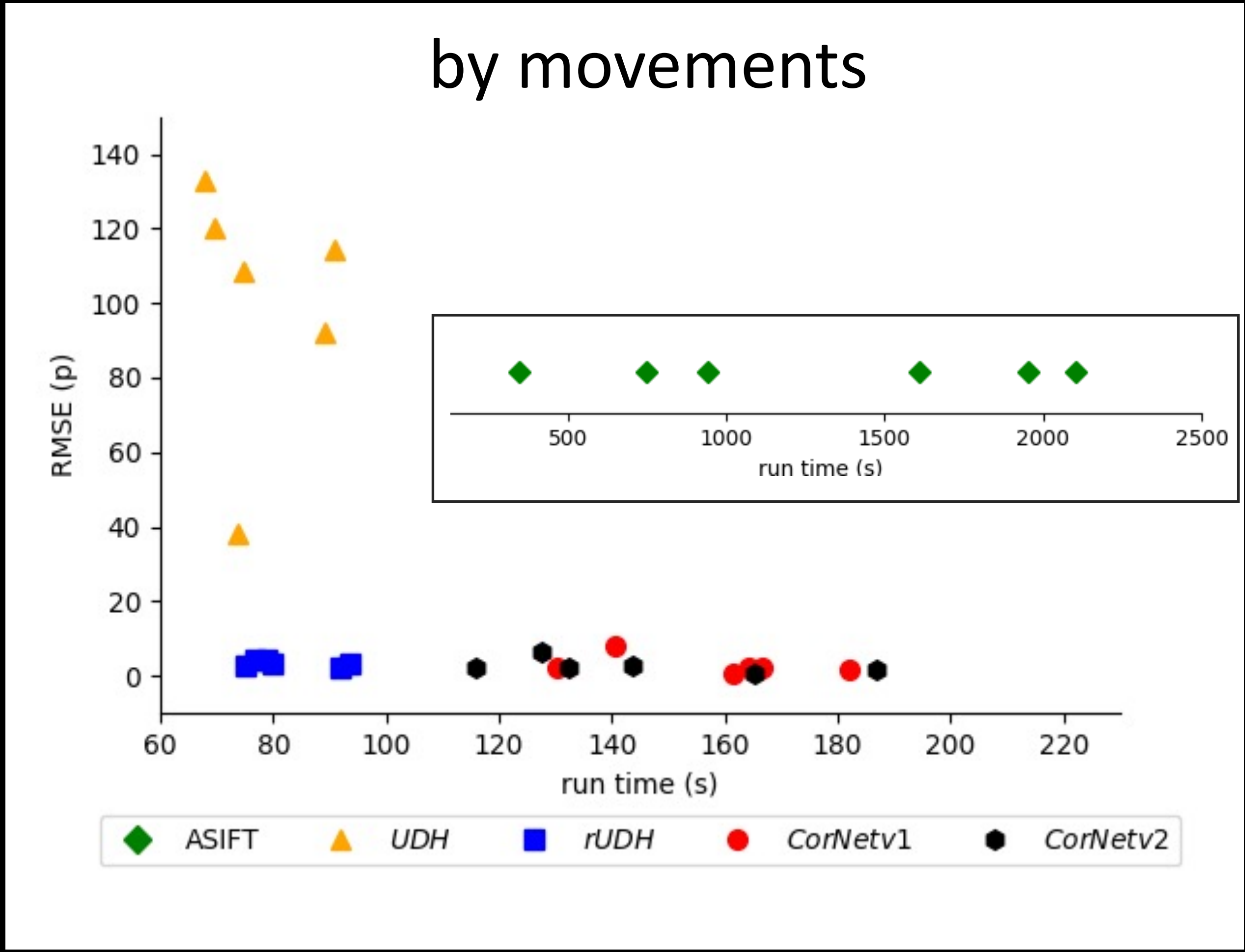
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Pred_i - GroundTruth_i)^2}{n}}$$

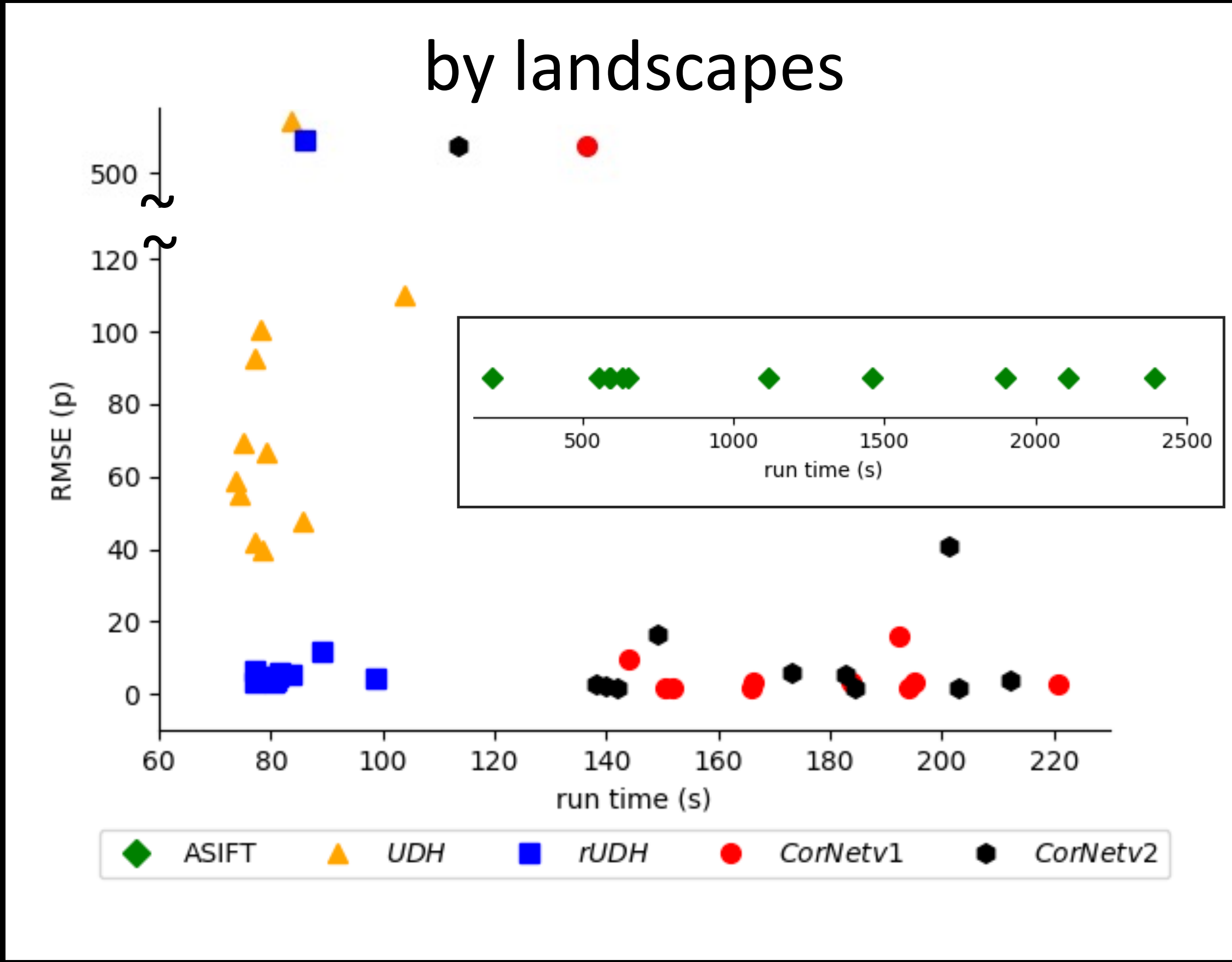
$$Distortion = \frac{Area\ of\ mosaic\ field}{Area\ of\ ground\ truth}$$



Ground truth	ASIFT	UDH	rUDH	*CorNet*	*CorNetv2*
RMSE (pixel)	-	74	2.7	2.3	2.4
Distortion	0.9613	0.7034	0.8428	0.9169	0.9623

Benchmarking





CorNetv1 vs CorNetv2 Comparison

$CorNetv1: Farm\ Buildings$

$CorNetv2:\ Farm\ buildings$

$CorNetv1: Fall\ Forest$

$CorNetv2: Fall\ Forest$

Future Work

- More detailed comparison between *CorNetv1* and *CorNetv2*
- Better handling of oblique camera poses
- Compare *CorNetv2* to other SOTA deep learning feature extraction algorithms such as LF-Net, SuperPoint, and D2-Net
- Dynamic sampling of frames: using Farneback dense optical flow to get a more uniform overlap between frames
- Lens and gimbal correction from an object inside the frame instead of a checkerboard
- Mini mosaicking algorithm to minimize error accumulation

Acknowledgments

- NSF Midwest Big Data Hub Digital Agriculture Spoke
- Missouri Maize Center
- National Corn Growers Association
- U. S. Army Research Laboratory grant W911NF-1820285
- An anonymous gift
- Hadi Aliakbarpour, Shizeng Zhou, Behirah Hartranft, Steven Suddarth, Madeline Hartranft, and Vinny for helpful discussions