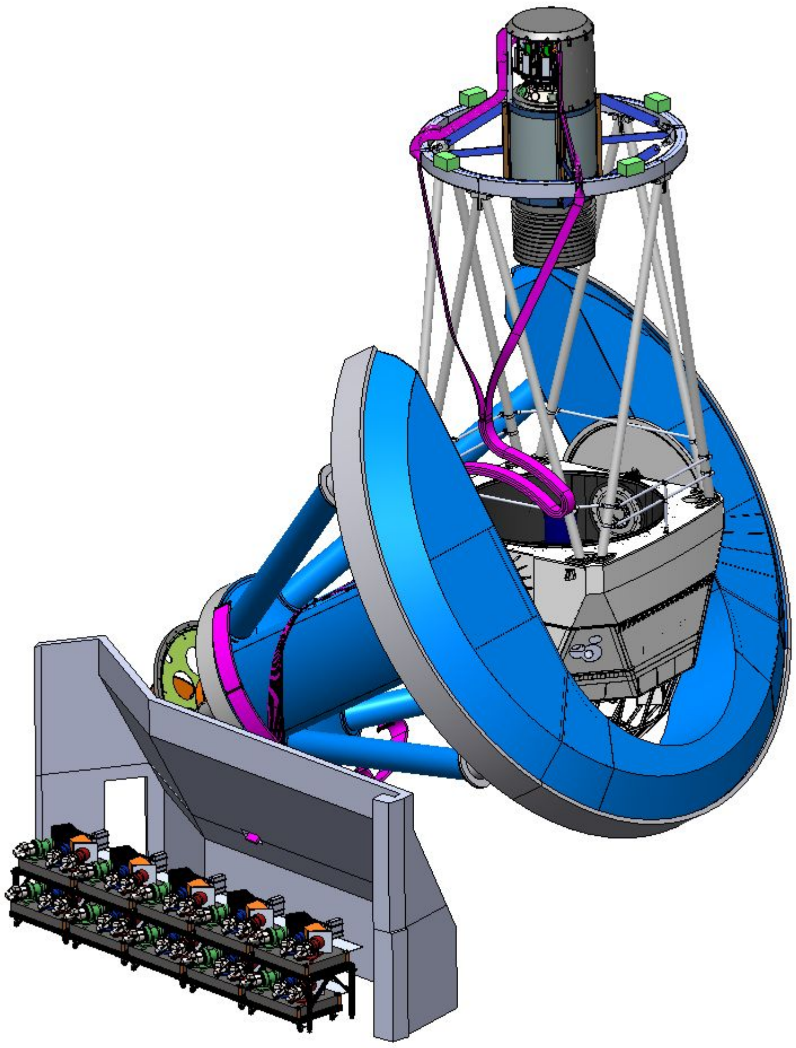


Deep Learning for Cosmic Ray Identification

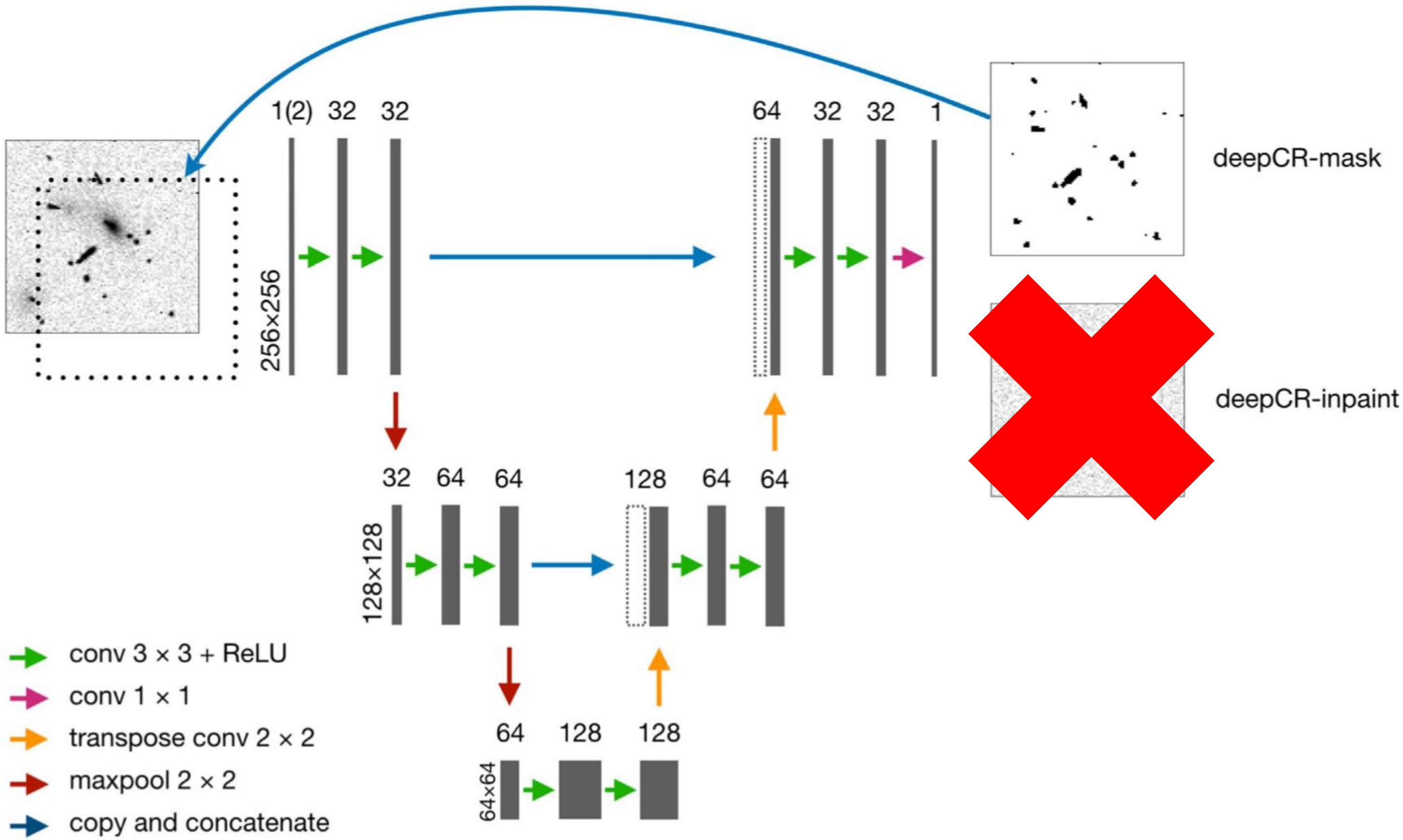
Dylan Green, UC Irvine
David Kirkby, UC Irvine
DESI-6095

Motivation

- SDSS method is “good” but not the best
  - Relies heavily on tweaking quixotic parameters for optimal results
    - Nsig, cfudge, c2fudge
  - Dependent on exposure PSFs
  - Can sometimes miss cosmics that are dim
  - Can easily overestimate cosmic footprint
- Fully Convolutional Neural Network (CNN) could alleviate previous problems by being trained to compensate for features like fibers
  - Only one tunable parameter of obvious intent (probability threshold)
  - “Black box” nature is a downside
- Deep learning for cosmics has been successfully applied on images taken on the HST

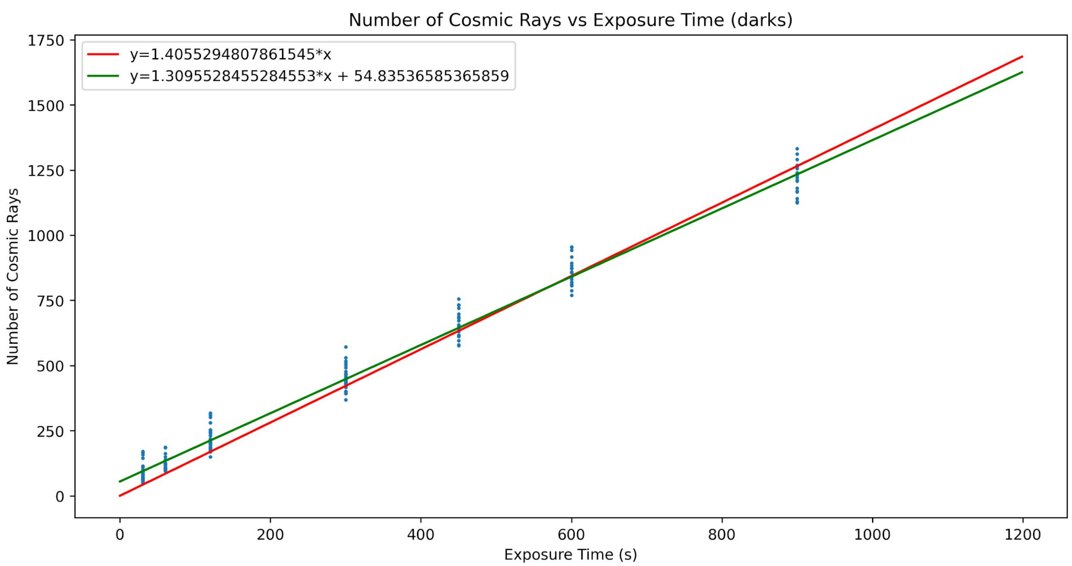
Model

- CNN architecture defined in deepCR (Zhang & Bloom 2020)



Data

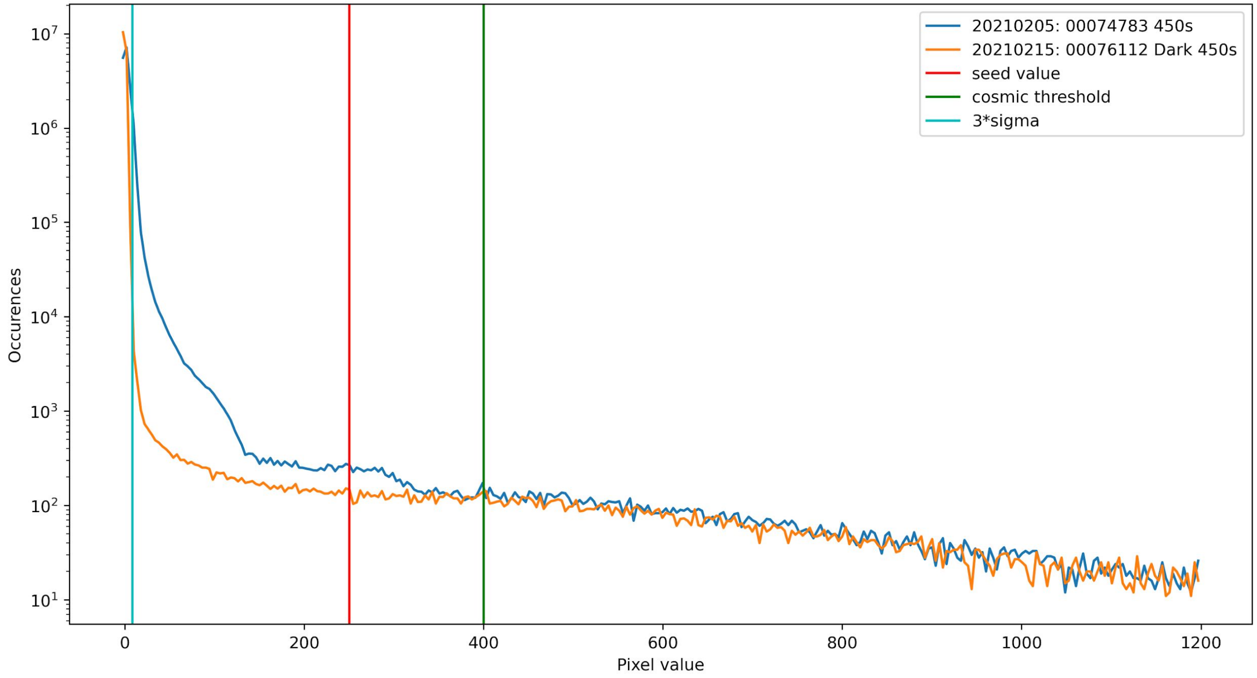
- Generate training images using “clean” images
  - Median of multiple exposures of identical tiles
  - Currently using R-band exposures from SV0 (3/14/20-3/15/20)
- Simulated noise added to training images
  - Noise picked from Gaussian with sigma ~sqrt(1/median_ivar) and mu~0
- Cosmics from 5 minute calibration darks 01/13/2021-01/21/2021
  - ~ 41k unique cosmics



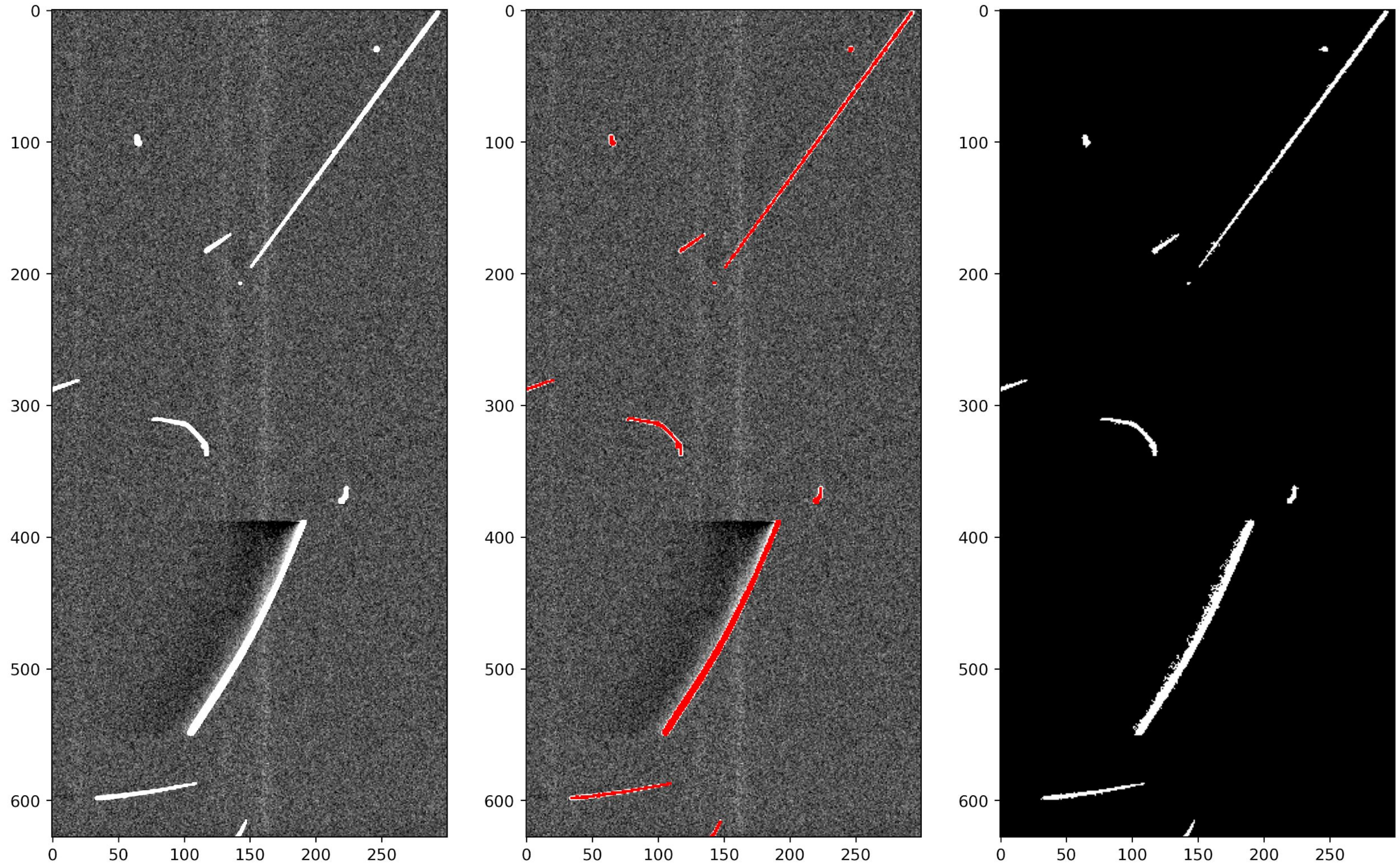
- Simulate number of cosmics via simulating exposure times
  - Approx. linear relationship between exposure time and num of cosmics
  - Num cosmics = ~1.4*exposure time

Extracting Cosmics

- Cosmics are extracted from darks
  - From histogram analysis everything above pixel value ~400 is a cosmic ray
- Can use a similarly high value (250) as a “seed” for cosmics
- Flood fill outwards from seed values until pixels drop below 3 sigma of a clean dark (sigma~1-3)

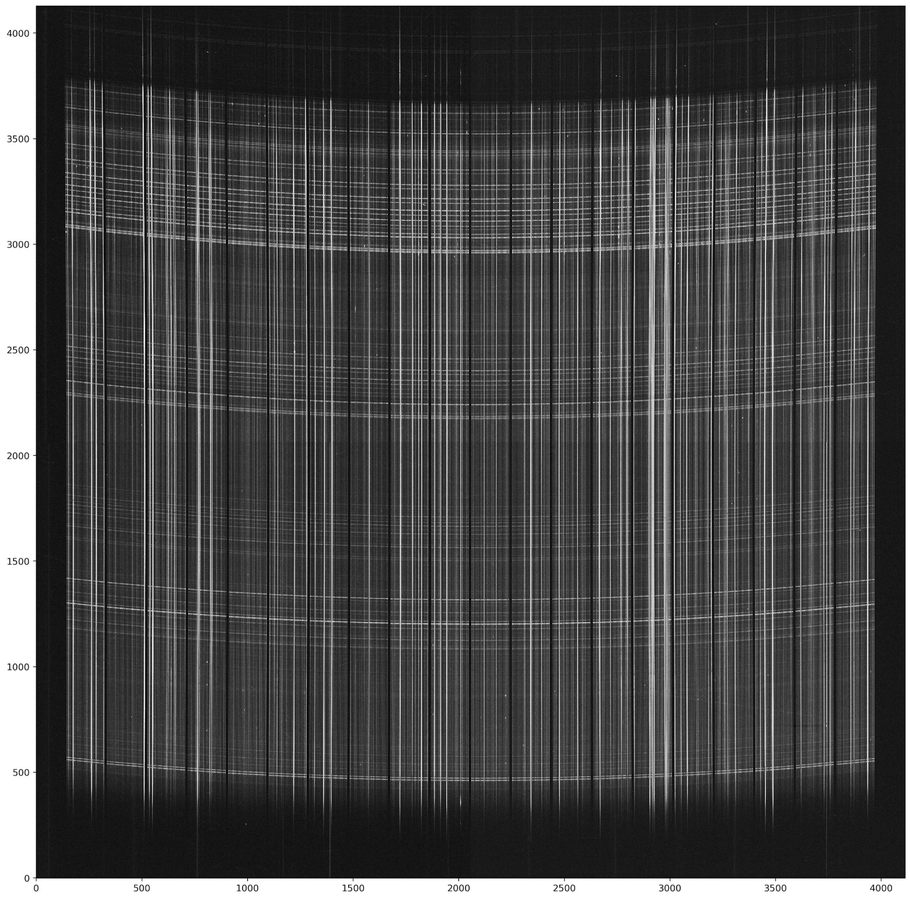


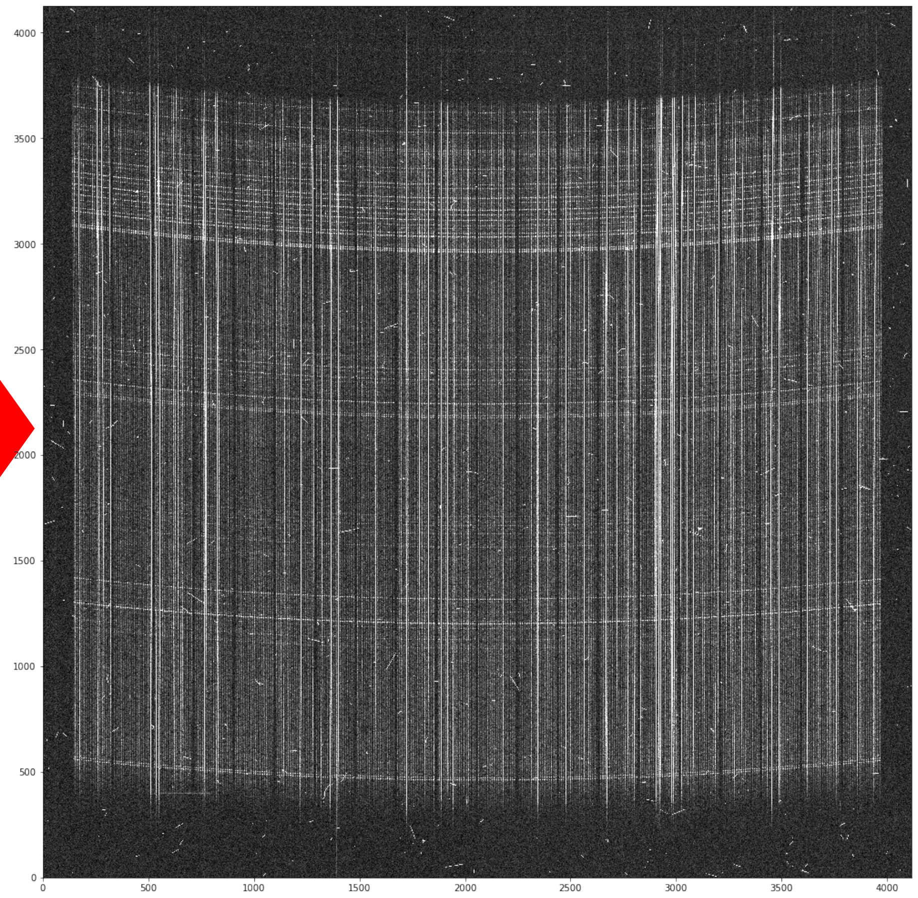
Extracting Cosmics



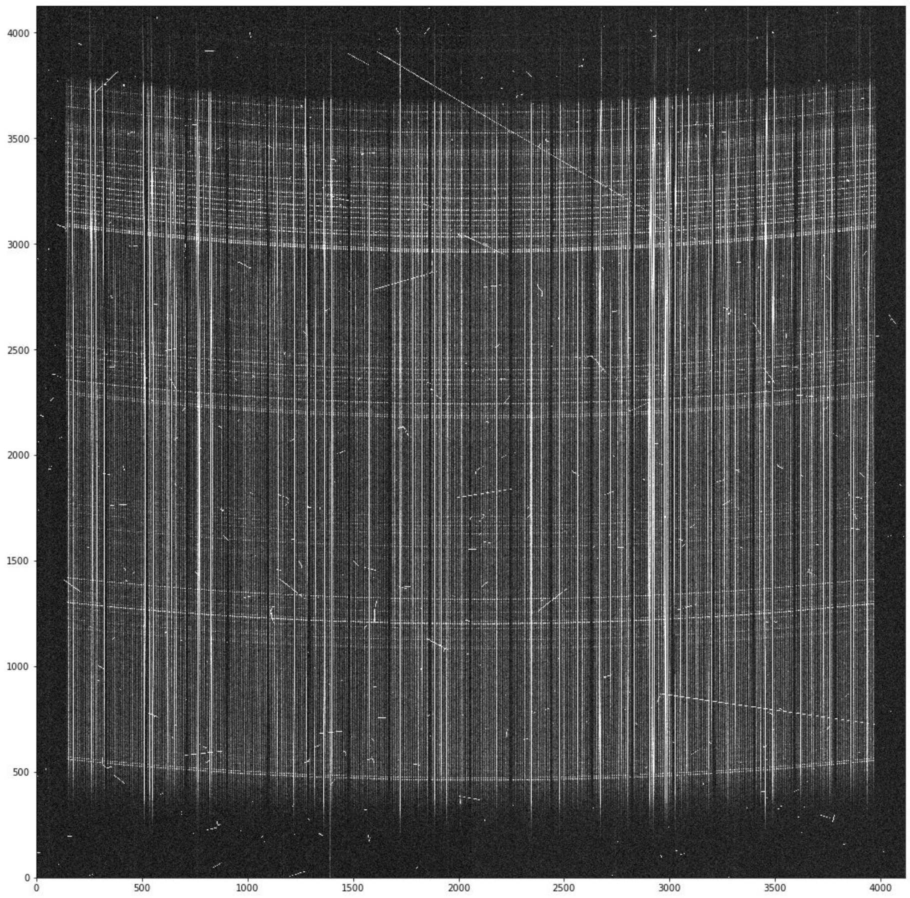
Data

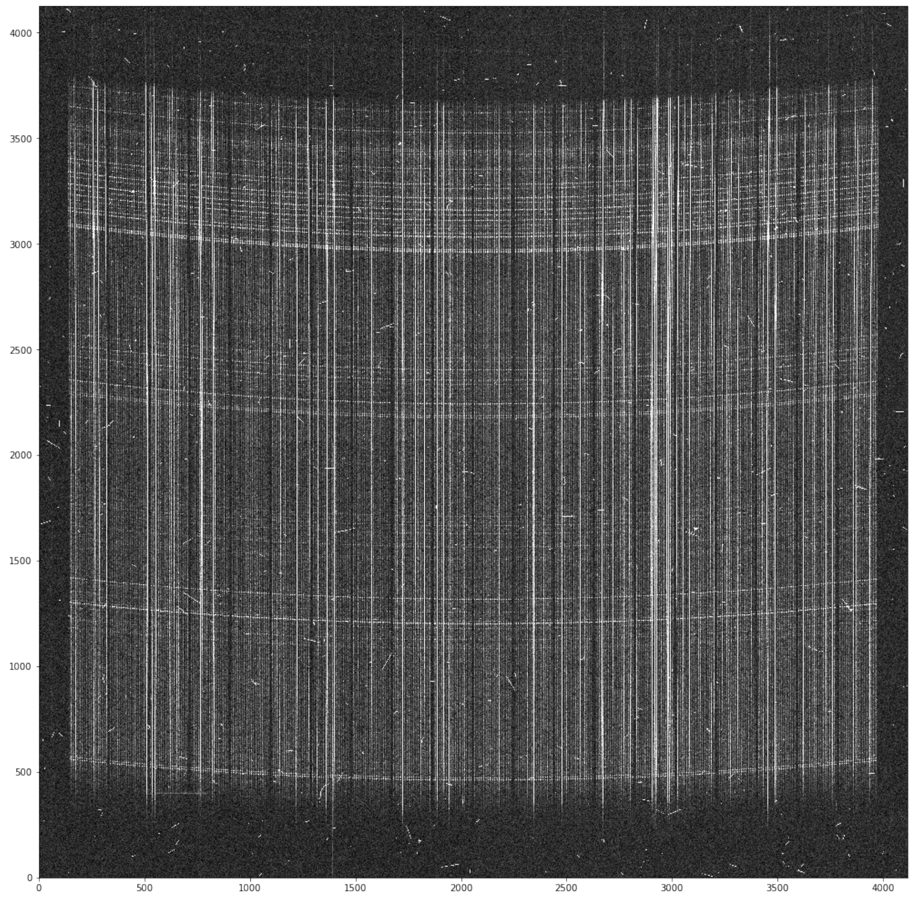
Median of Tile ID #66003





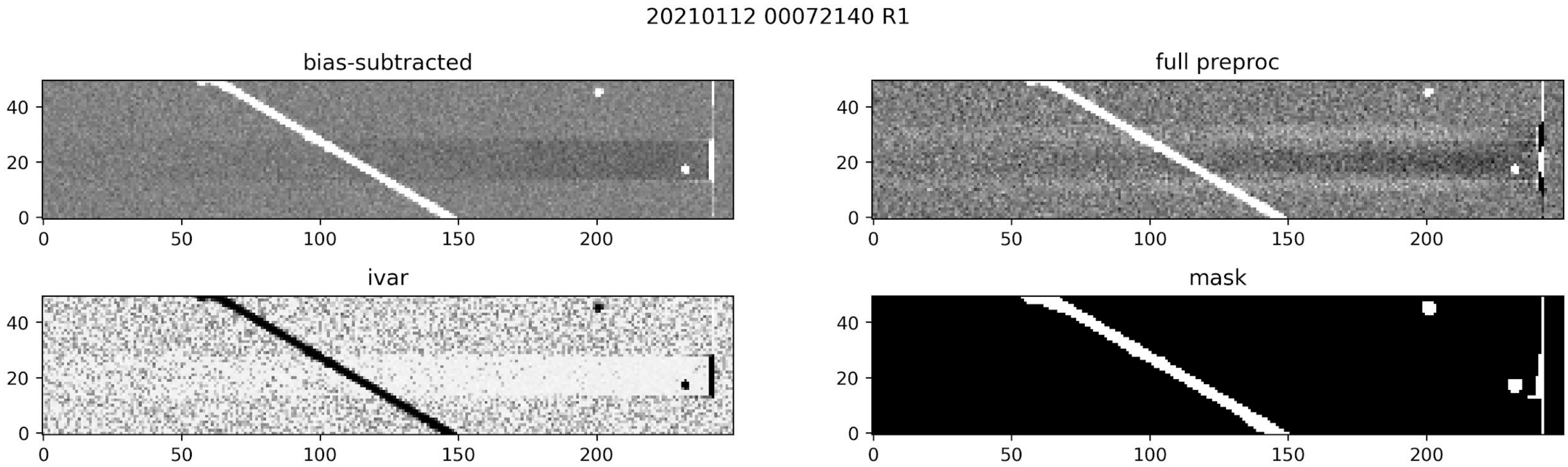
Data


4000
3500
3000
2500
2000
1500
1000
500
0
0
500
1000
1500
2000
2500
3000
3500
4000


4000
3500
3000
2500
2000
1500
1000
500
0
0
500
1000
1500
2000
2500
3000
3500
4000

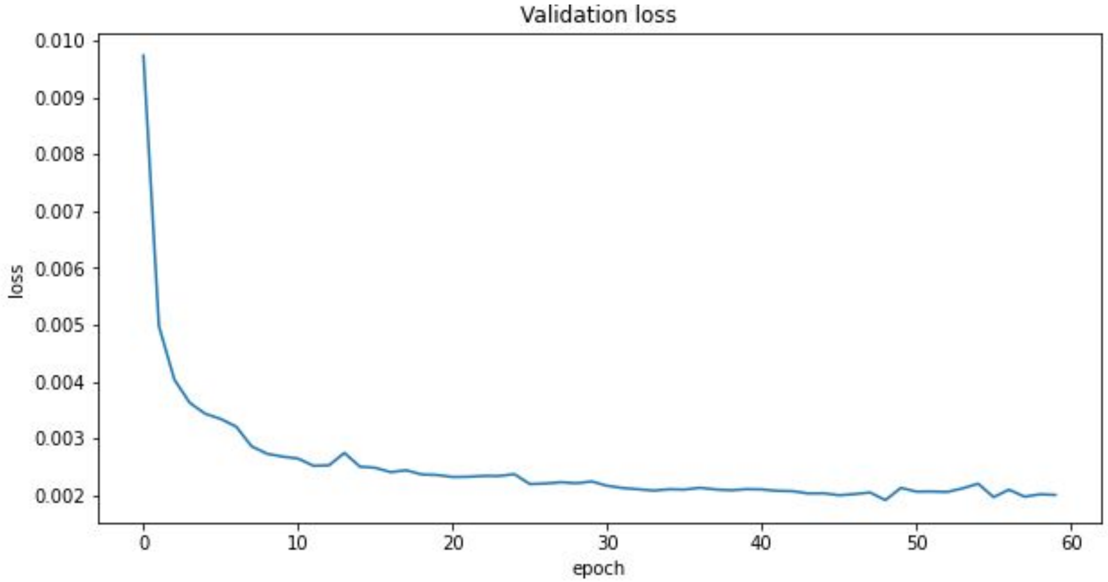
Dark Trails

- In current preproc pipeline dark trails are corrected before identifying cosmics
- Training data does not correct dark trails (yet)
- More important for SDSS algorithm which requires ivar
  - CNN does not use ivar

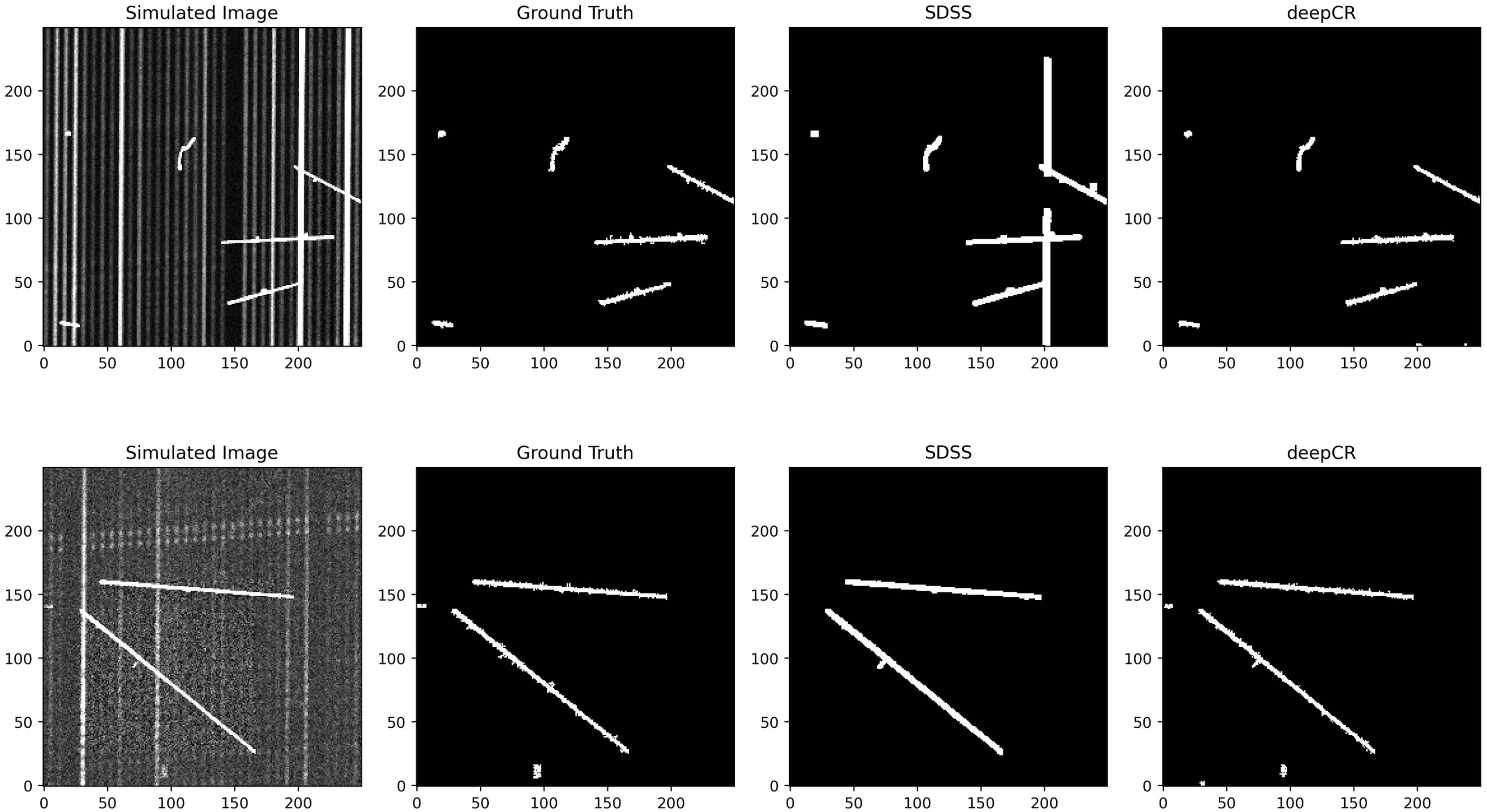


Training

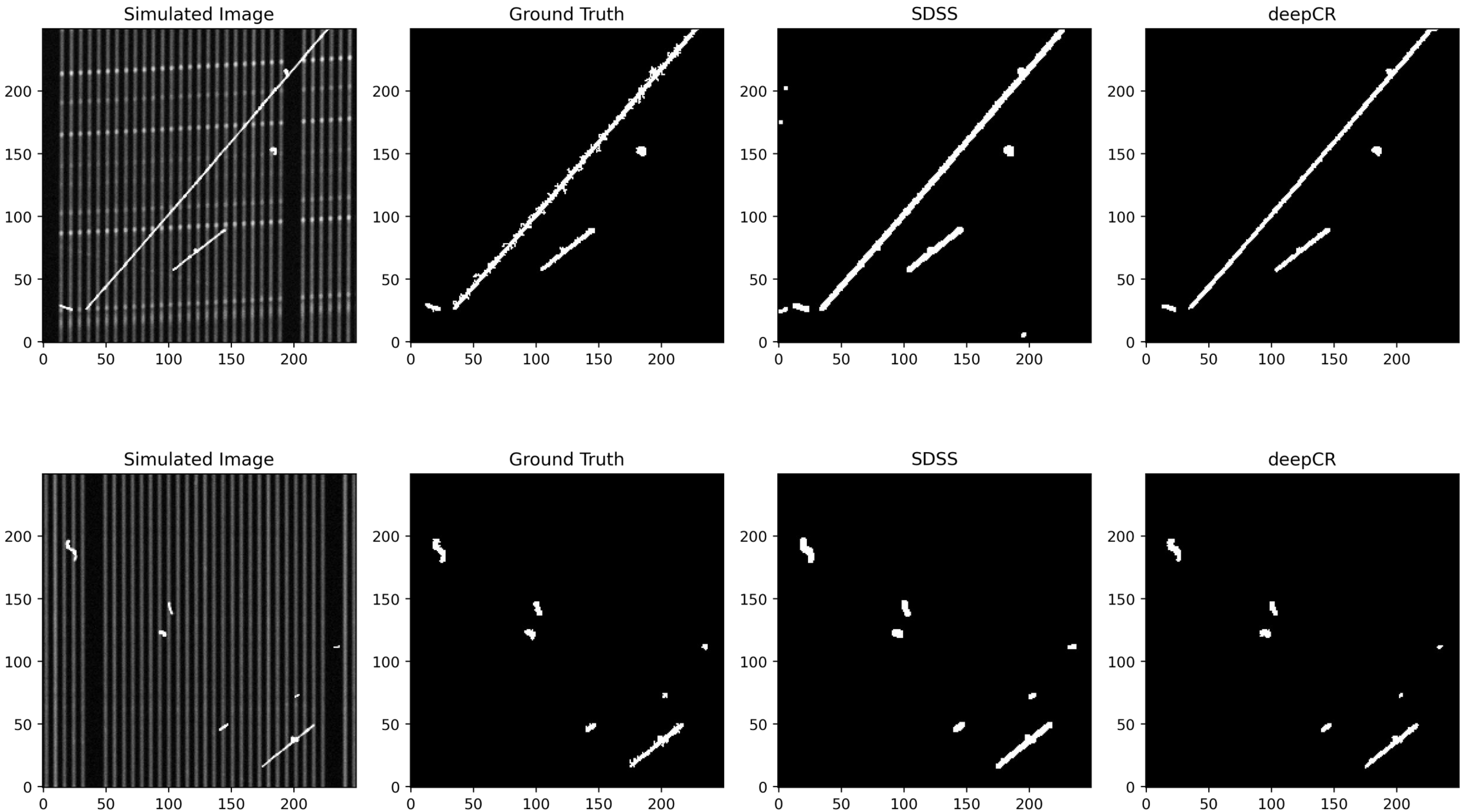
- Images are cut into overlapping 256x256 patches
  - 10 training images -> 2404 patches, 1803 (75%) of which are used for training
- Network outputs “probability map”
  - > 0.5 probability threshold identified as a cosmic
  - Tunable parameter
- Train for 60 epochs
  - Batch normalization statistics frozen after 24 epochs



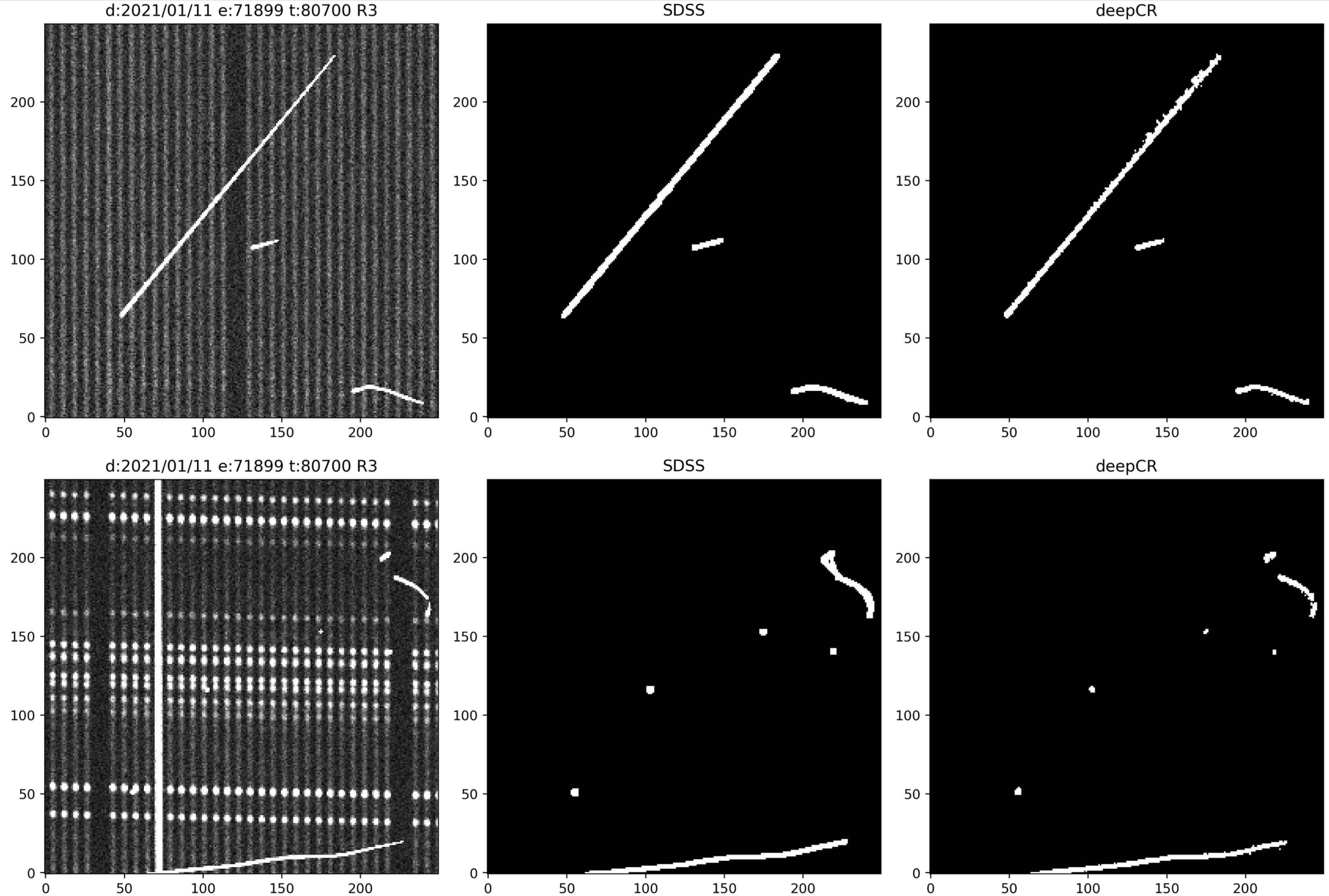
Comparison (Simulated Data)



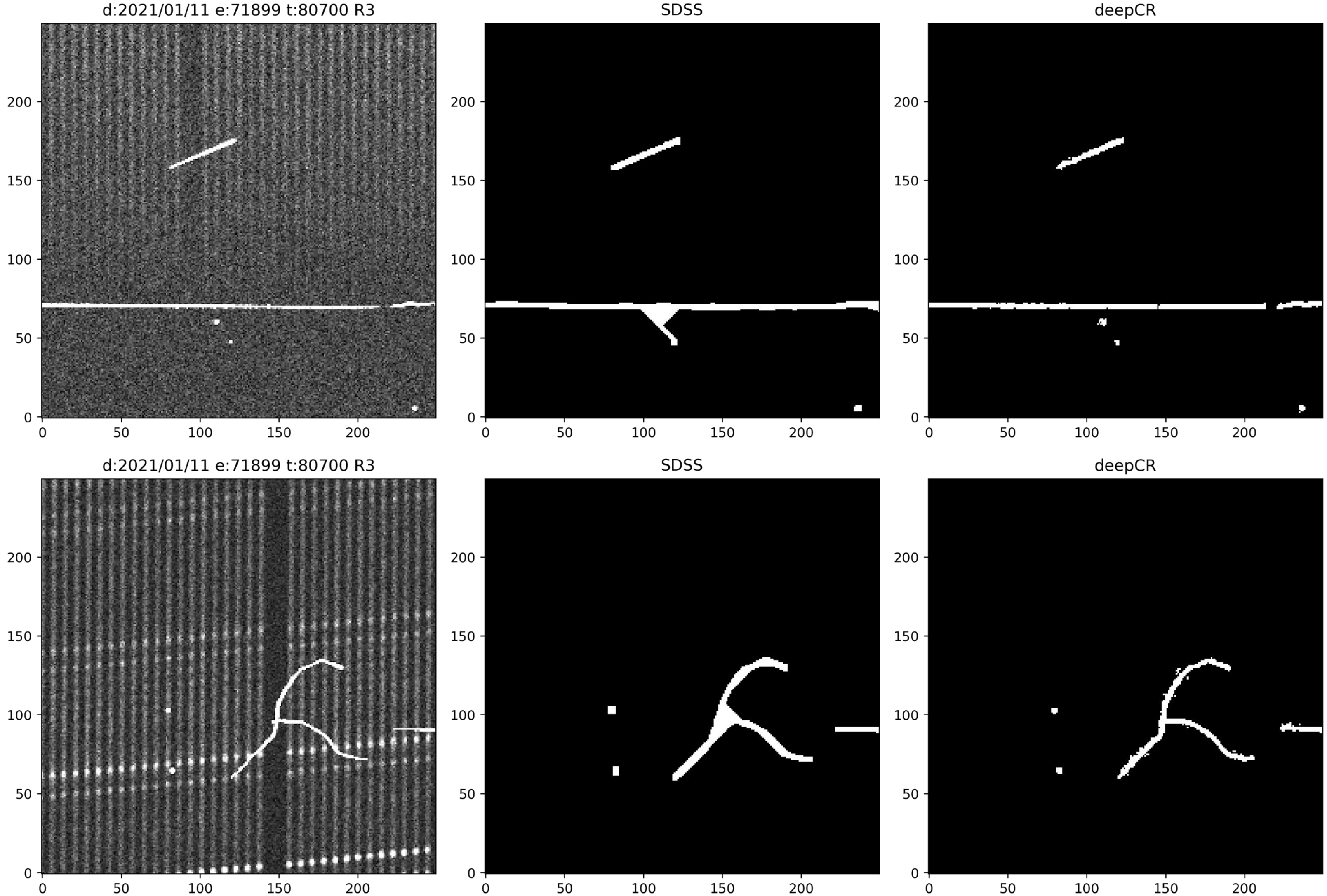
Comparison (Simulated Data)



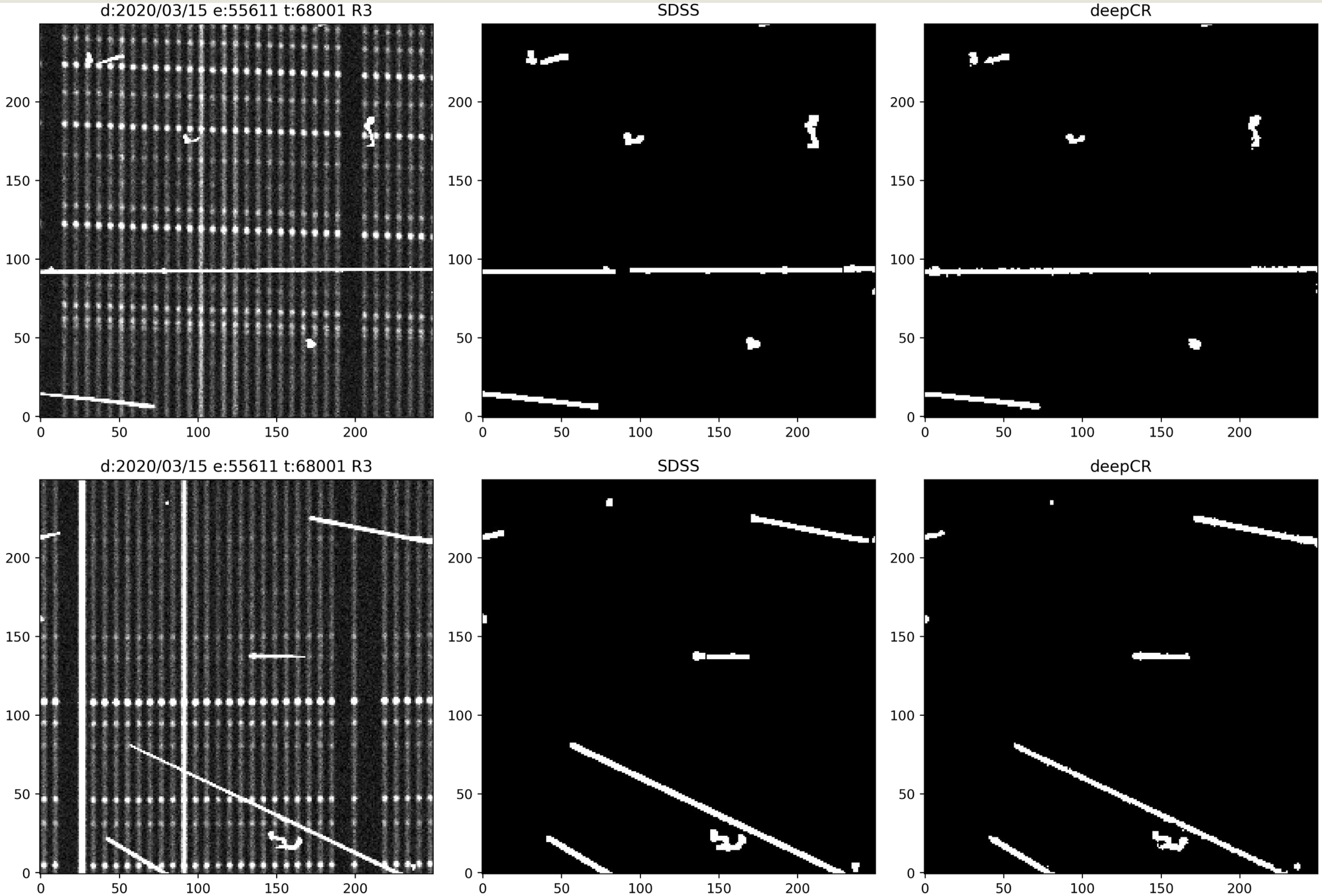
Comparison (Real Data 2021/01/11)



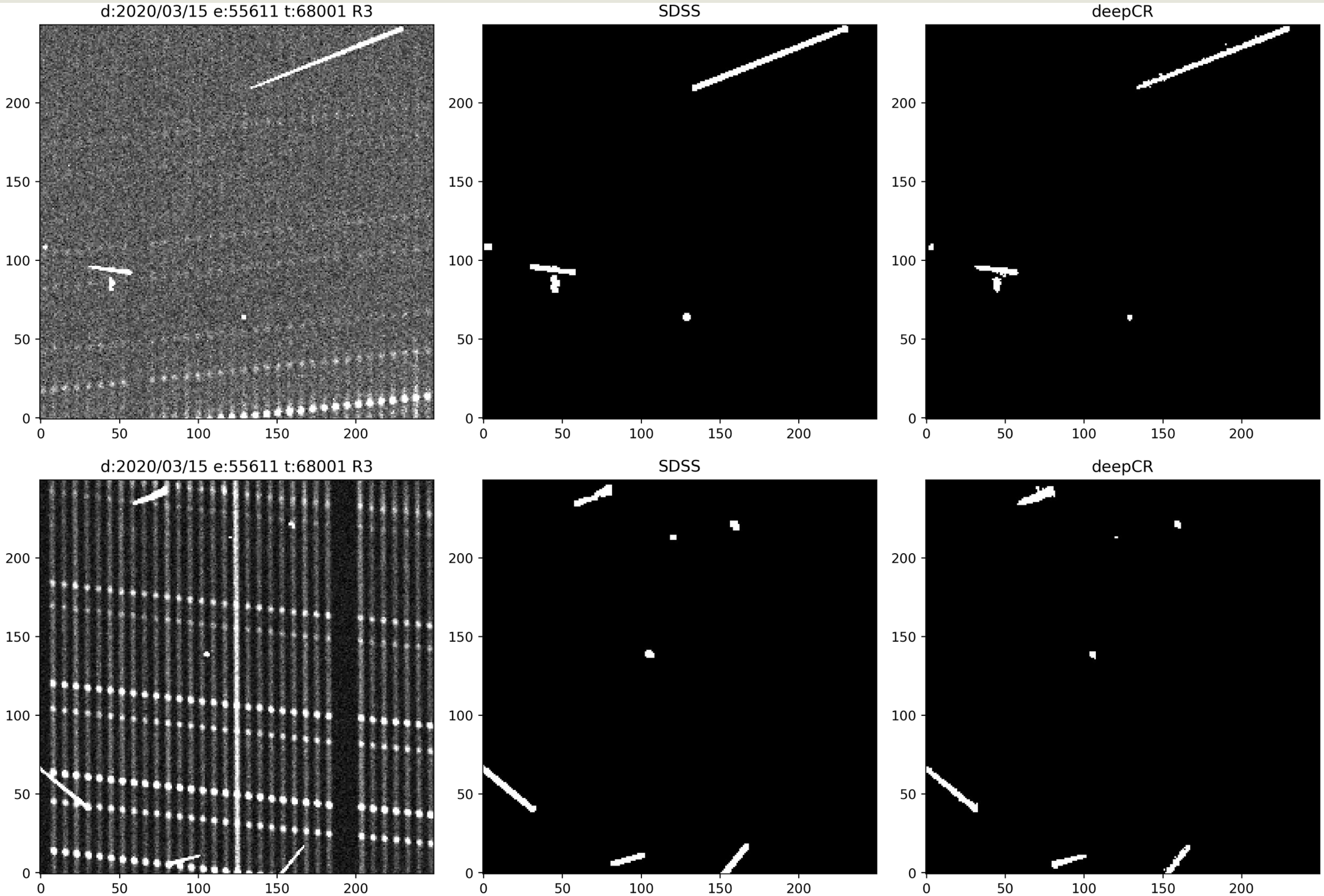
Comparison (Real Data 2021/01/11)



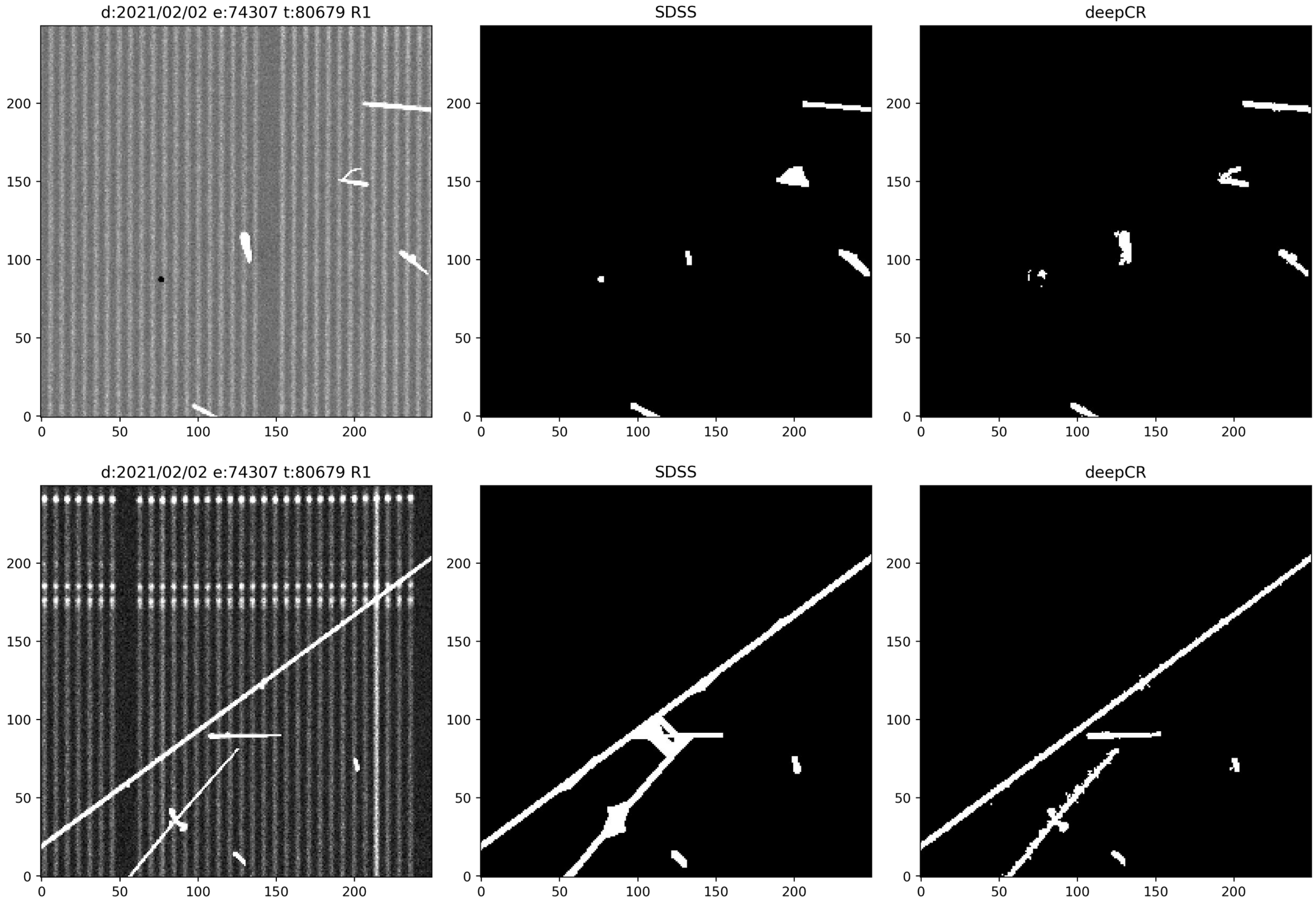
Comparison (Real Data 2020/03/15)



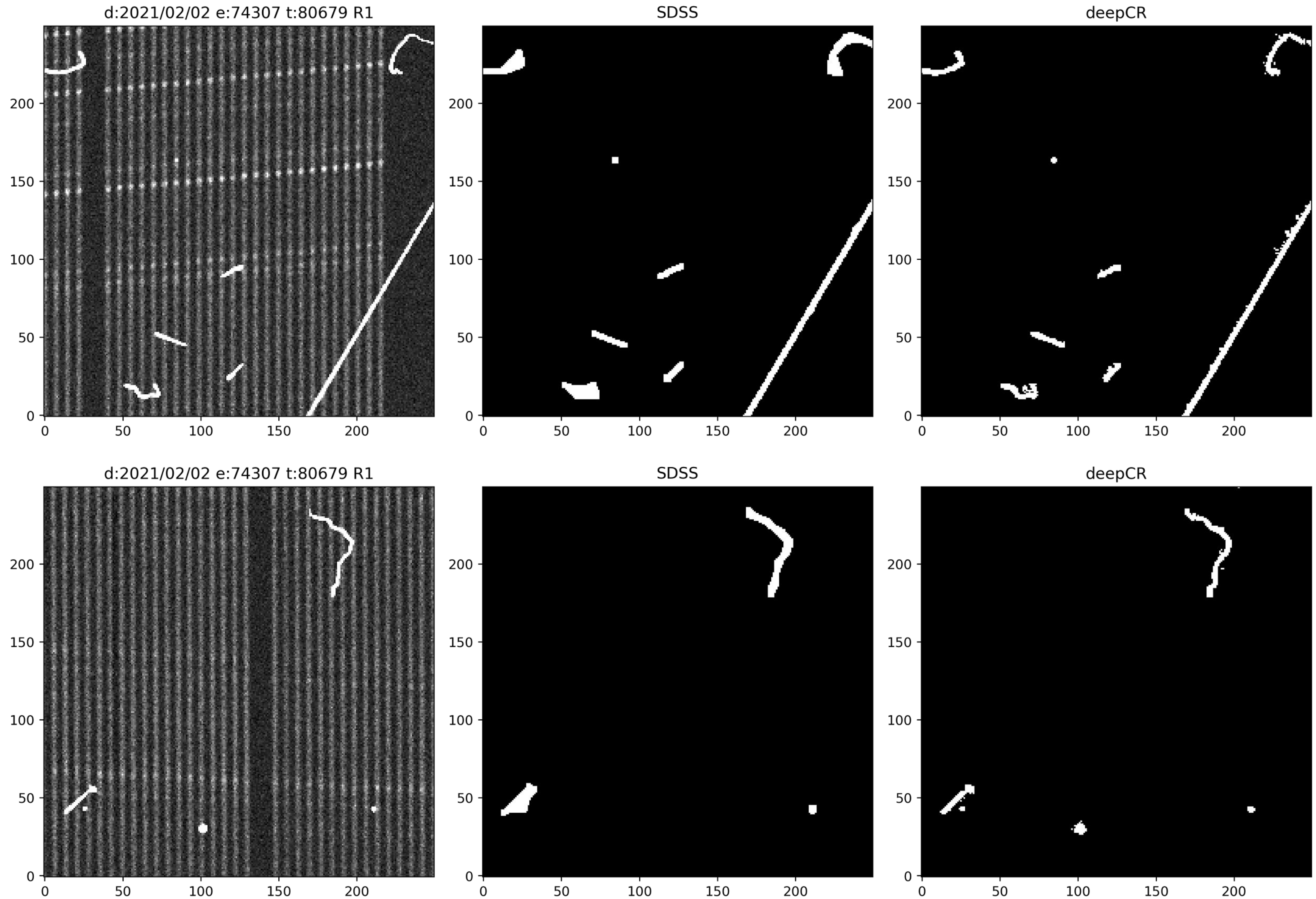
Comparison (Real Data 2020/03/15)



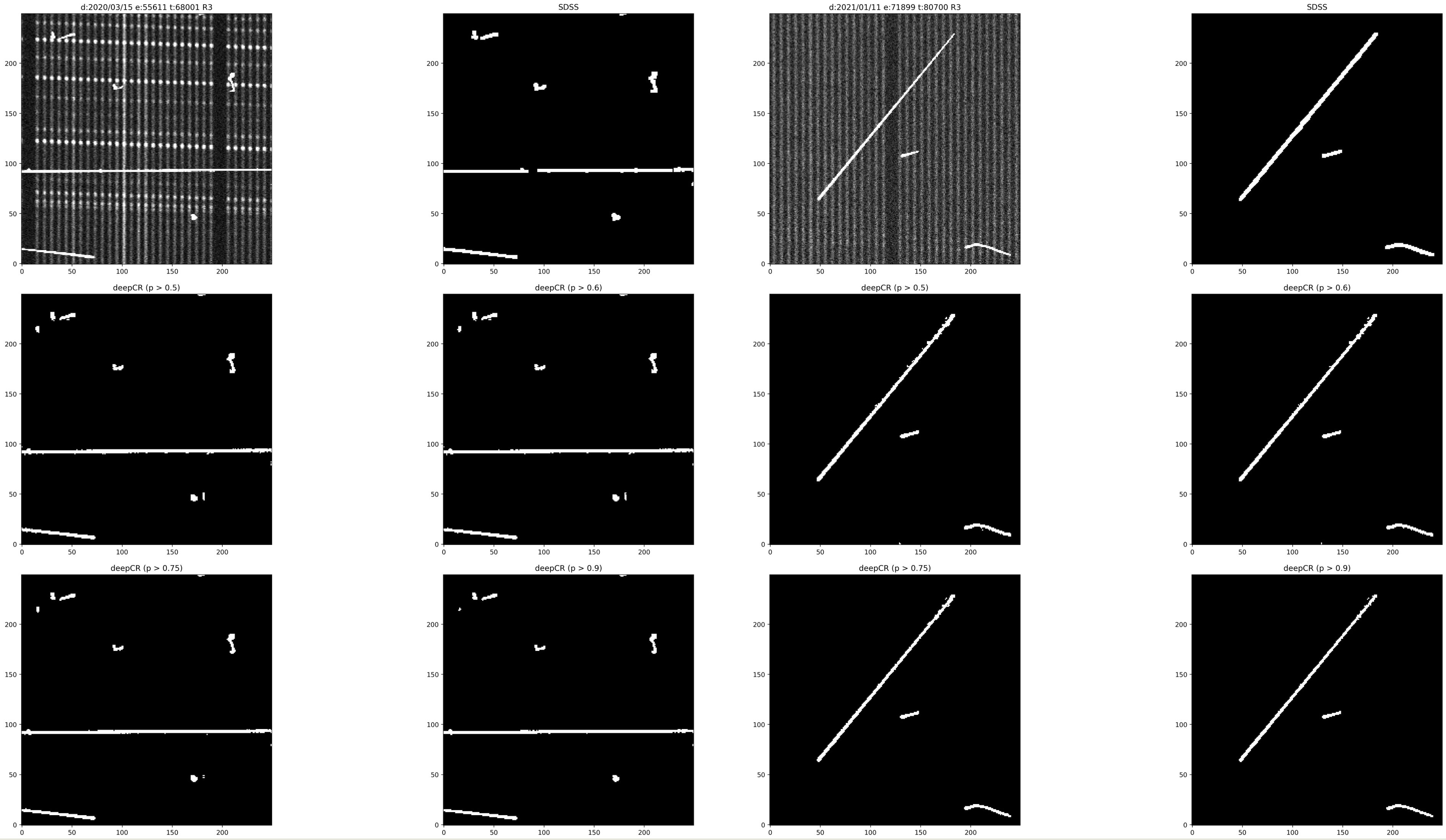
Comparison (Real Data 2021/02/02)



Comparison (Real Data 2021/02/02)



Comparison (Real Data)



Timing

- ~5s for deepCR impl on low end (NVIDIA Tesla P4) GPU machine
  - Large memory requirement (~fewGB) to process 4k x 4k image at once.
  - Reduced to ~2-4s if we process the image in 16 smaller quadrants
  - Training ~20 minutes
- About 2x as fast on NERSC GPU node
  - NERSC won't give me enough memory to run the whole image :(
  - ~1-1.5s if we process the image in 16 smaller quadrants
- ~45s on my laptop (when nothing else is running)
  - Reduced to ~30s if we process the image in 16 smaller quadrants
  - Training ~40 minutes
- ~5-25s (depends on number of cosmics) for SDSS

Next Steps

- Train individual weights for B and Z band.
- Improve training data to match more how SDSS algorithm “sees” images
- Try and extract spectral data from exposures with cosmic rays marked by deepCR
- Build network independent of deepCR
  - Alternate network configuration may perform better on spectroscopic data
  - Smaller network may have similar performance while being faster
  - Larger network may not sacrifice speed for better cosmic ID