

A grid of lion faces used for data augmentation. The grid consists of 20 small images of lion heads, arranged in 4 rows and 5 columns. The lions are shown in various poses and orientations, illustrating the concept of data augmentation in machine learning.

Data Augmentation Crash Course

Dr. Wendy Hawley Carande



Laboratory for Atmospheric and Space Physics



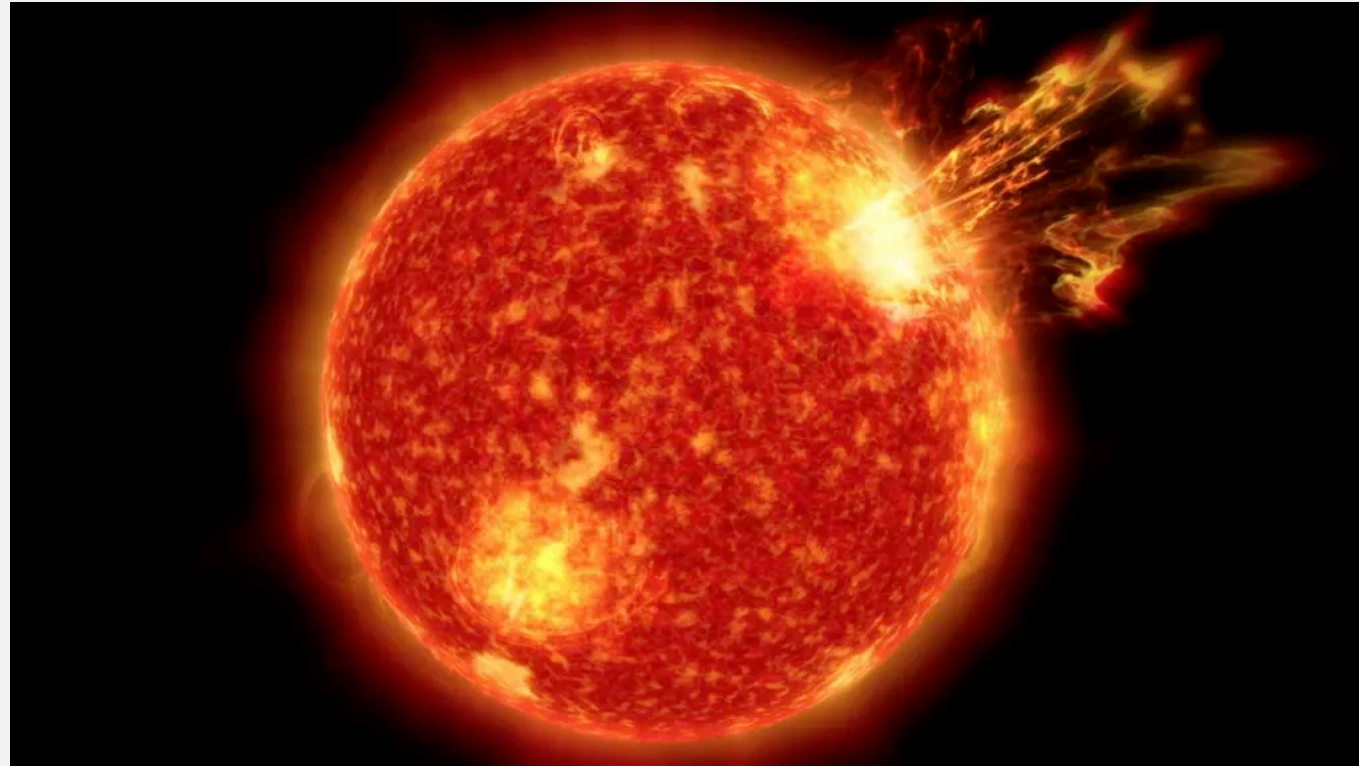
ABOUT ME



INSIGHT



BACKGROUND



Use machine learning to predict whether an solar flare will occur within a prediction window

MOTIVATION

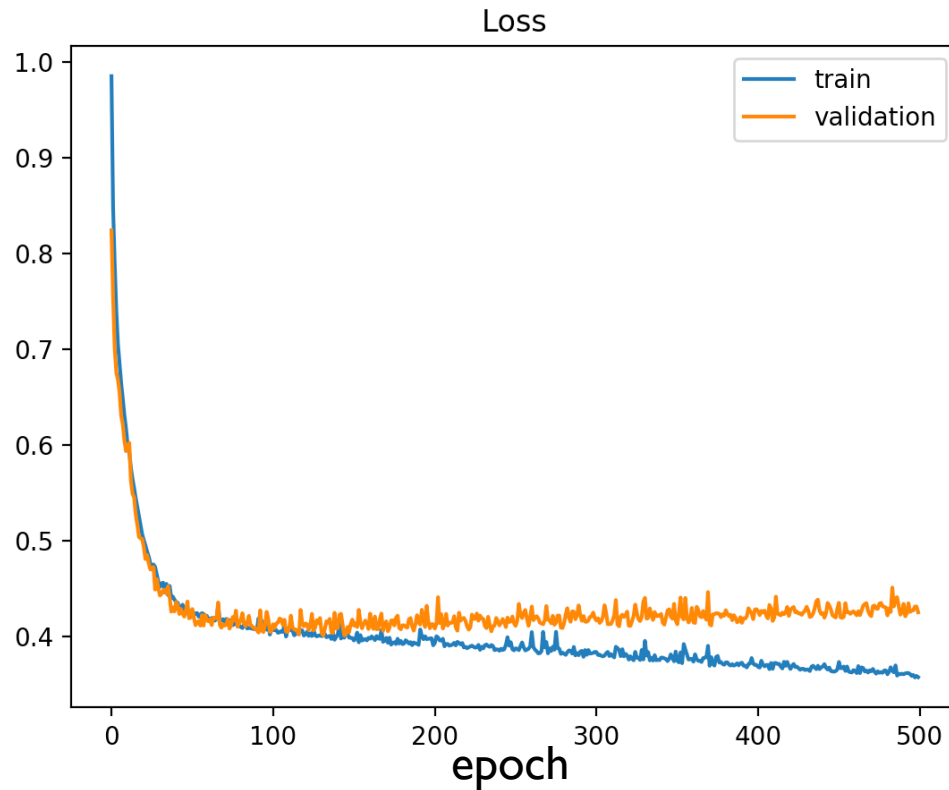


QUANTITY

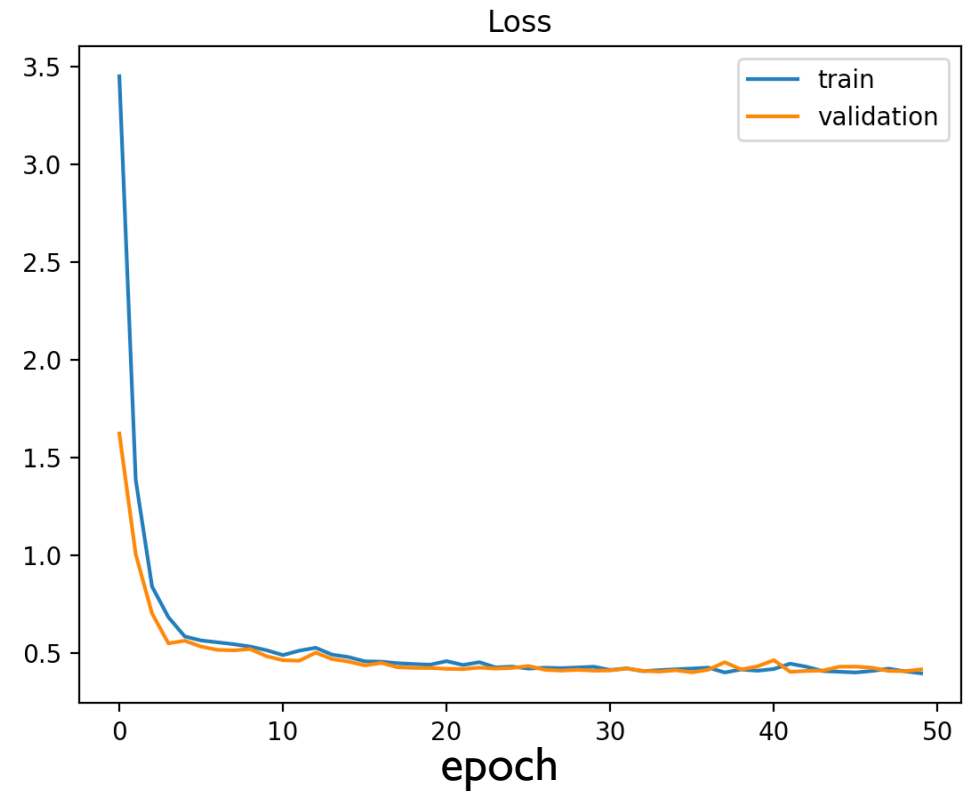
&

VARIETY

OVERFITTING



Overfit model



“Ideal” model

Dog



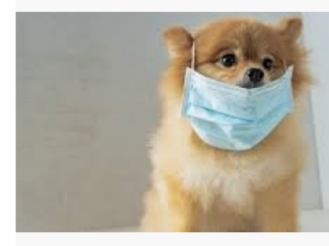
The Best Dogs Of BBC Earth | Top 5 ...
youtube.com



Dog - Wikipedia
en.wikipedia.org



How dogs contribute to your health ...
medicalnewstoday.com



Hong Kong dog causes panic – but her...
theconversation.com



convert dog years to human years ...
sciencemag.org



The 25 Cutest Dog Breeds - Most ...
goodhousekeeping.com



dog foods and heart disease
cnbc.com



How to Keep Your Dog Cool in the Sum...
pets.webmd.com



Southeastern Guide Do...
guidedogs.org



The dog did not develop symptoms of ...
livescience.com

VS.

Cat



Cats May Recognize Their Own ...
smithsonianmag.com



Thinking of getting a cat ...
icatcare.org



Coronavirus: Cat owners fear pets will ...
bbc.com



The Shots Your Cat Needs
pets.webmd.com



5 things that scare and stre...
timesofindia.indiatimes.com



Cat infected with COVID-19 from owner ...
livescience.com



His Cat's Death Left Him Heartbroken ...
nytimes.com



Doctor's Note: Cats and the coronavirus ...
aljazeera.com

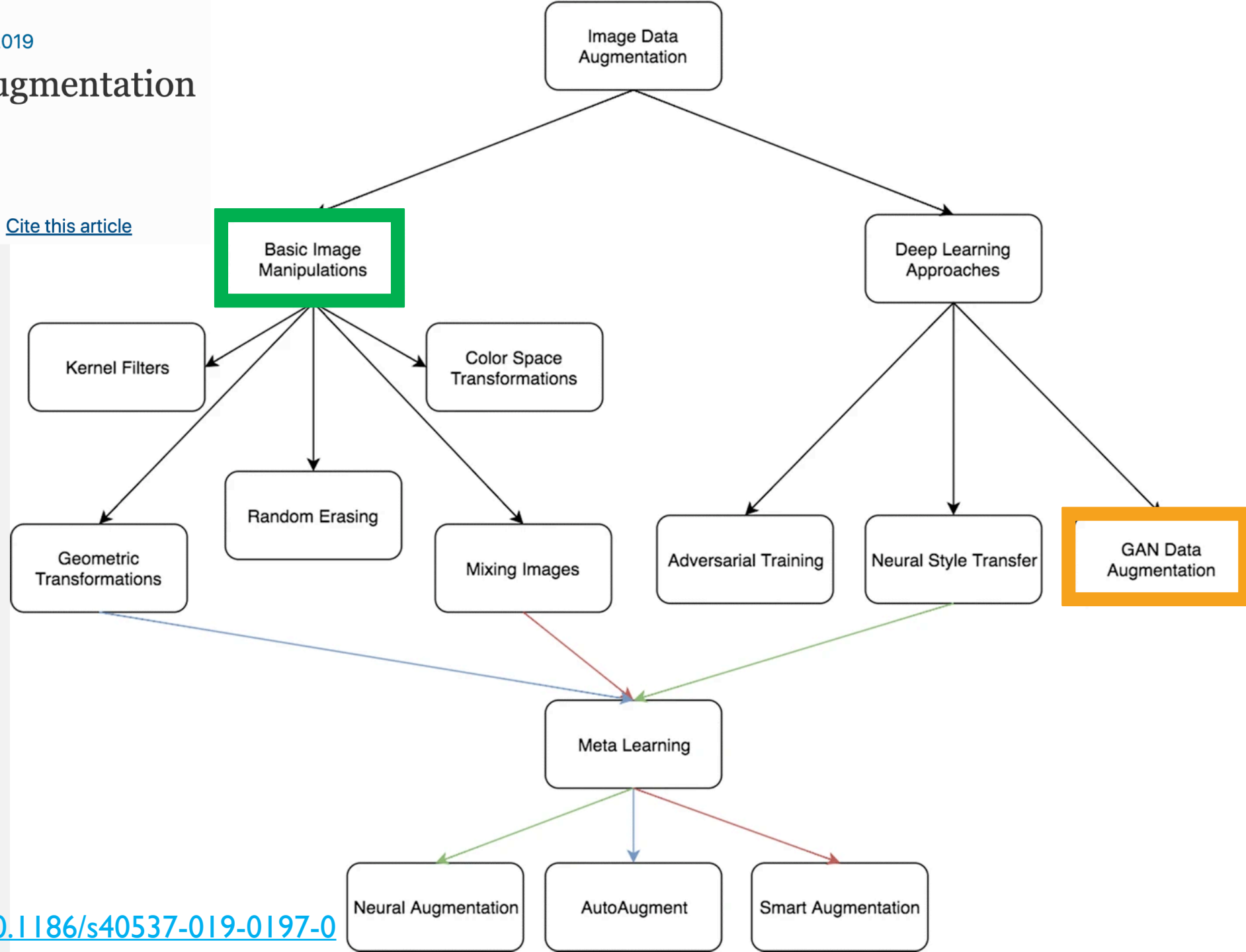


How to help a frightened cat | The ...
humansociety.org

A survey on Image Data Augmentation for Deep Learning

[Connor Shorten](#) & [Taghi M. Khoshgoftaar](#)

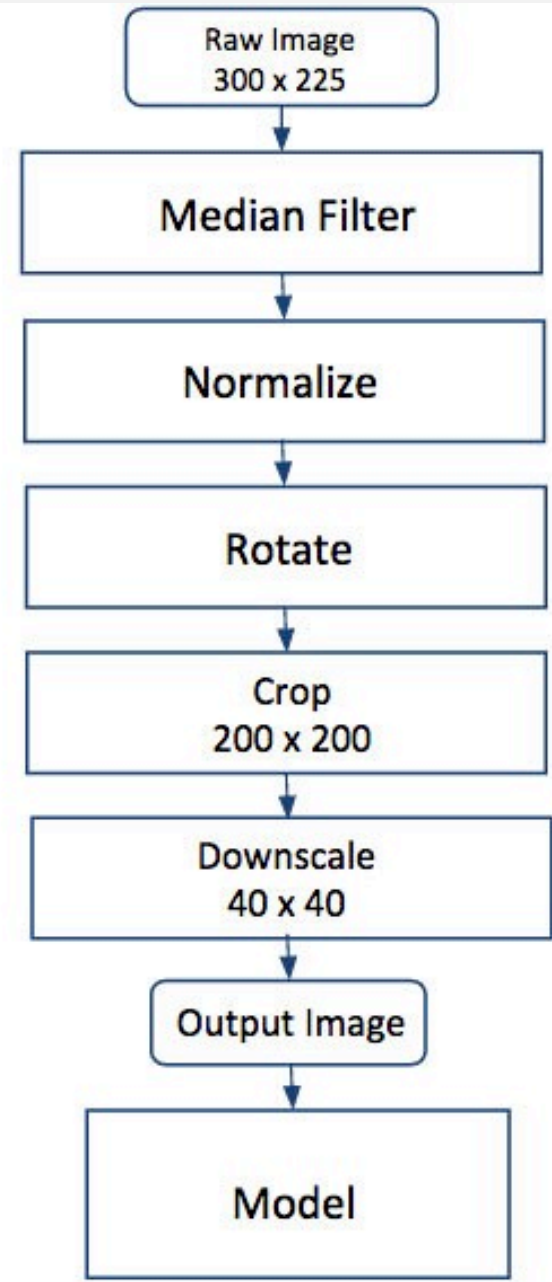
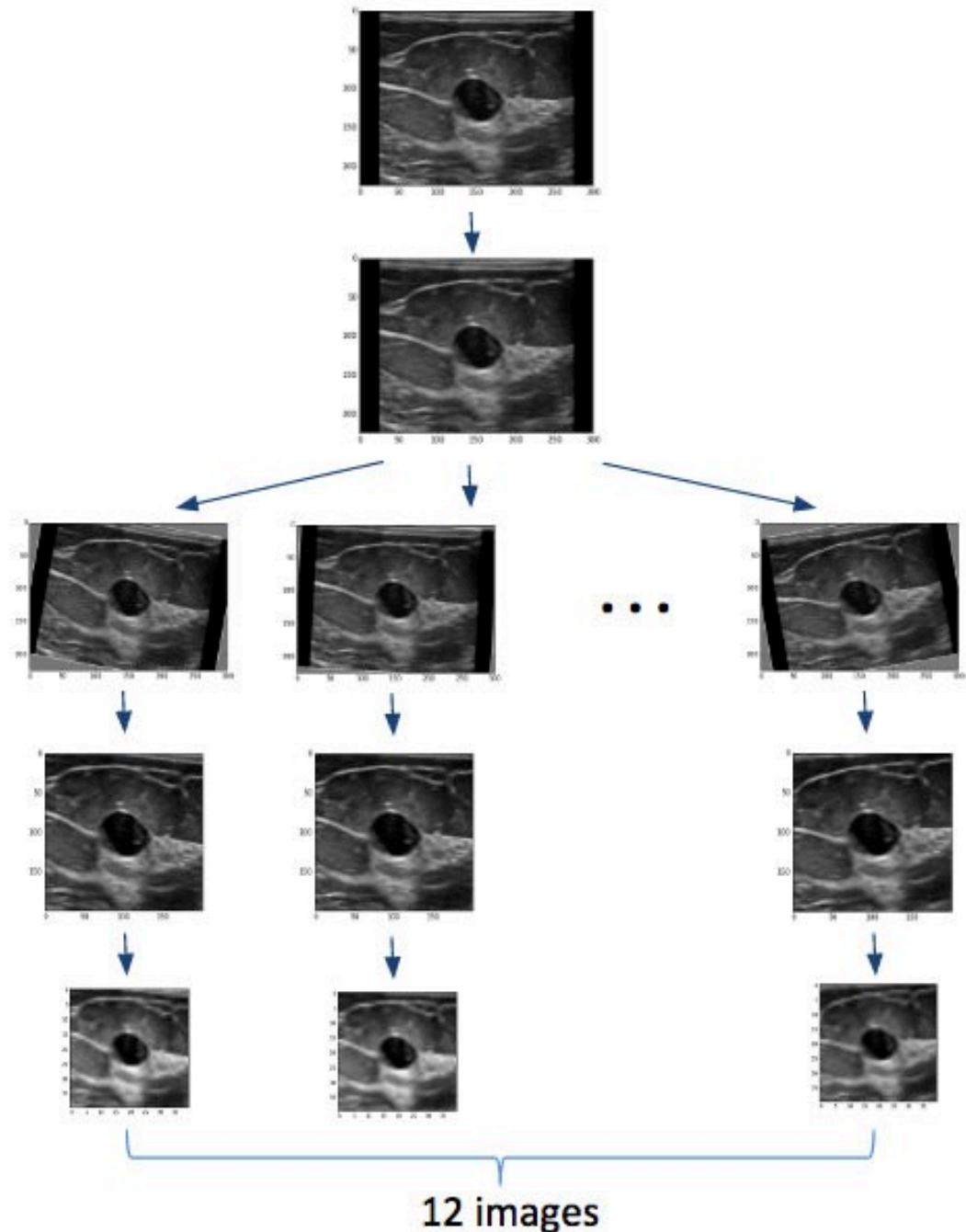
Journal of Big Data **6**, Article number: 60 (2019) | [Cite this article](#)



DEMO

ONE-LINE IMAGE TRANSFORMATIONS
IN SCIKIT-IMAGE

https://github.com/wcarande/data_augmentation/blob/master/augmentation.ipynb



DATA AUGMENTATION

- Use Generative Adversarial Networks to create synthetic magnetogram data
 - Trained/tested on an archive of SOHO/MDI data
 - Compatible with SDO/HMI magnetogram data
 - Validated on the overlap period between HMI and MDI