NCAR

2024 CESM AMWG Meeting

Wednesday, February 14, 2024

Informing Depositional Ice Growth Models Through 3-D Reconstruction of Ice Crystal Images Using Machine Learning

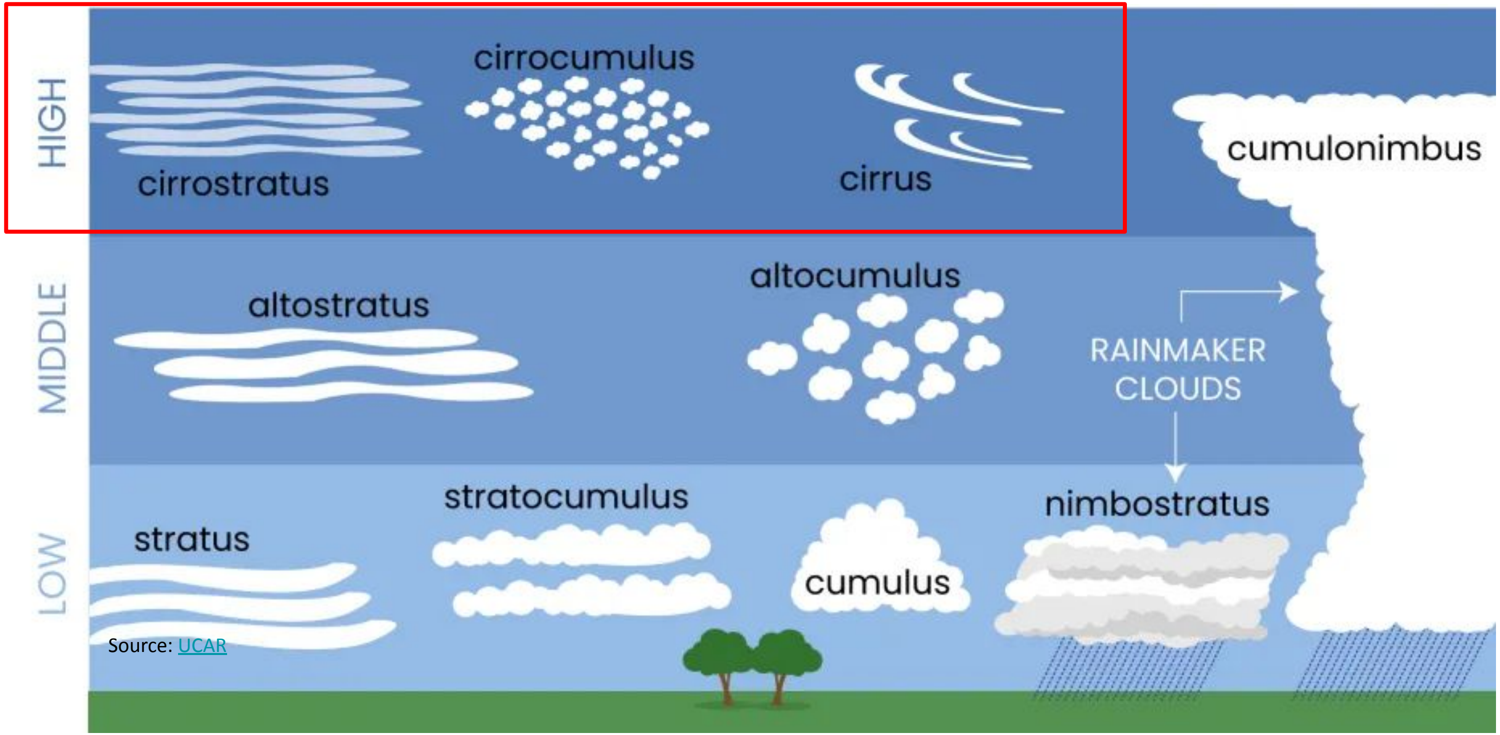
Joseph Ko[1], Jerry Harrington[2], Kara Sulia[3], Vanessa Przybylo[3], Marcus van Lier-Walqui[1,4], Kara Lamb[1]

[1] Columbia University, LEAP (Learning the Earth with AI & Physics) NSF STC
[2] Pennsylvania State University
[3] University at Albany
[4] NASA Goddard Institute for Space Studies

Source: climateandweather.net

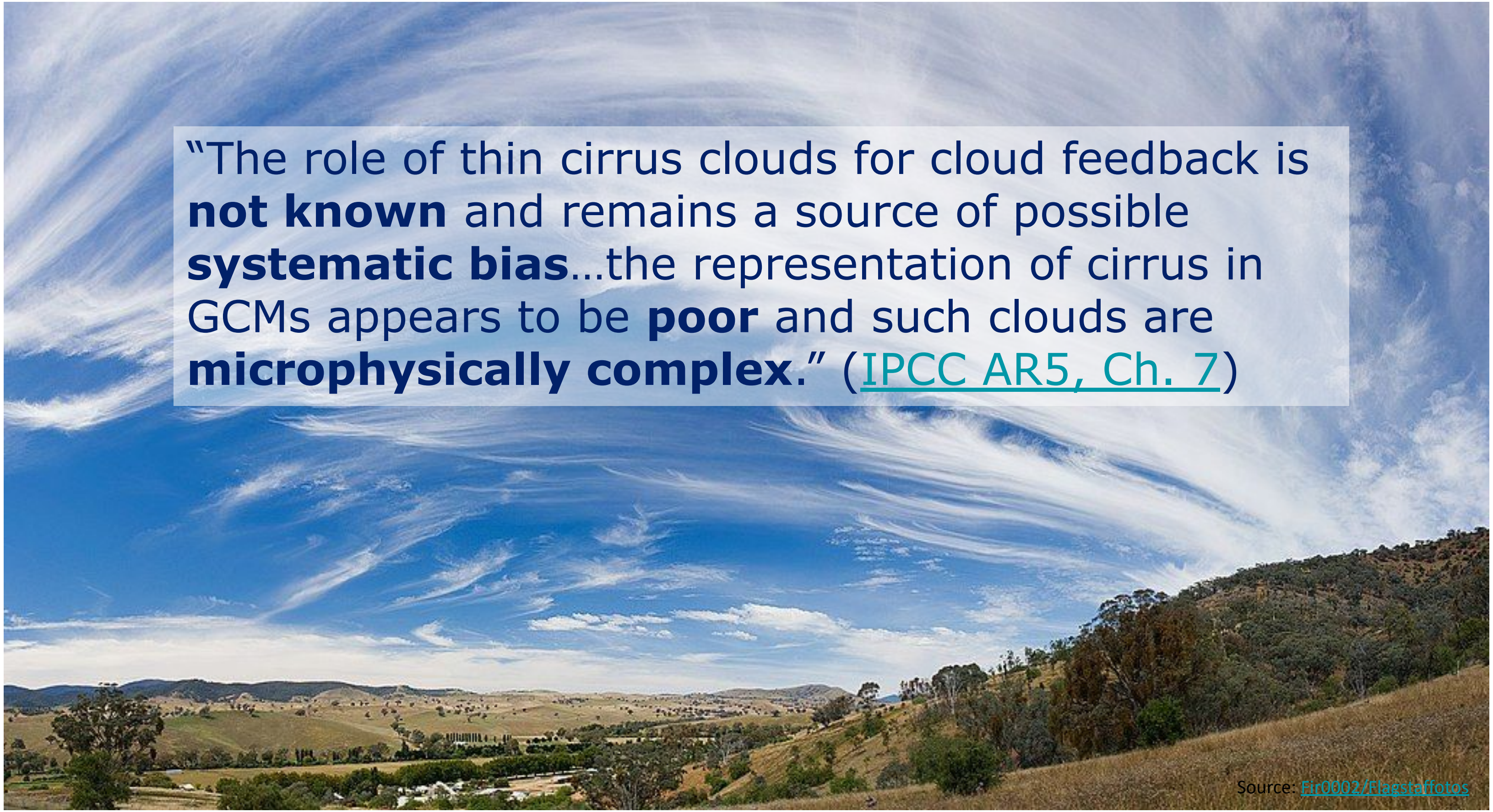
Columbia University in the City of New York | LEAP | NSF | PennState | University at Albany State University of New York | NASA

Clouds strongly impact the climate



Clouds impact Earth's **energy balance** and **hydrologic cycle**

Ice clouds are poorly understood

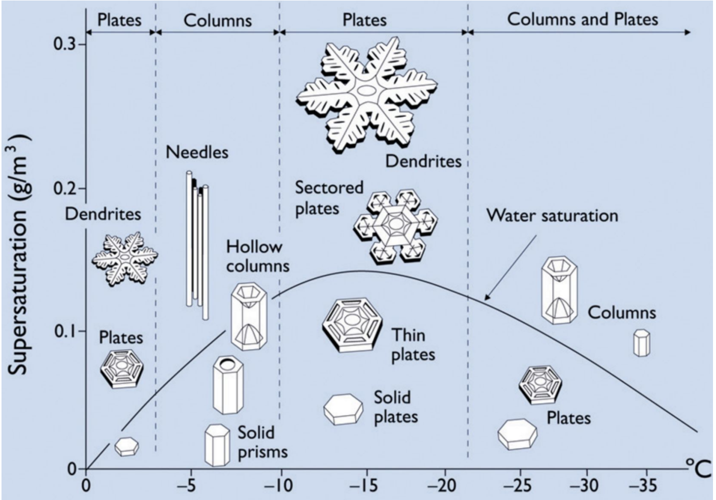


"The role of thin cirrus clouds for cloud feedback is **not known** and remains a source of possible **systematic bias**…the representation of cirrus in GCMs appears to be **poor** and such clouds are **microphysically complex**." (IPCC AR5, Ch. 7)

Source: Fir0002/Flagstaffotos

Ice crystal shape matters

Source: Kenneth Libbrecht, snowcrystals.com



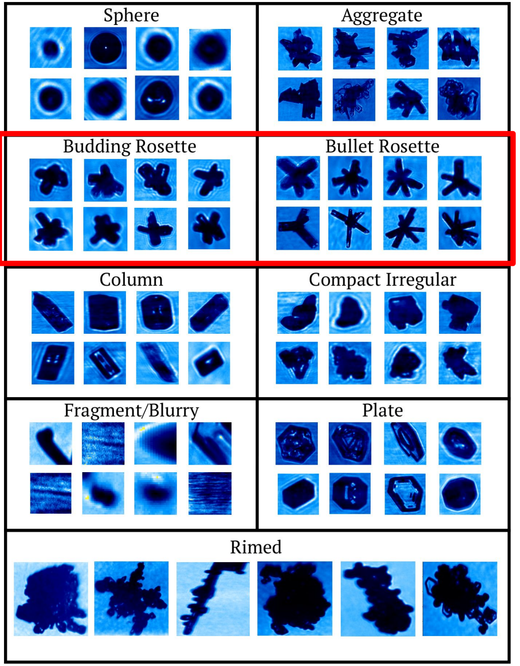
Source: Libbrecht, 2005

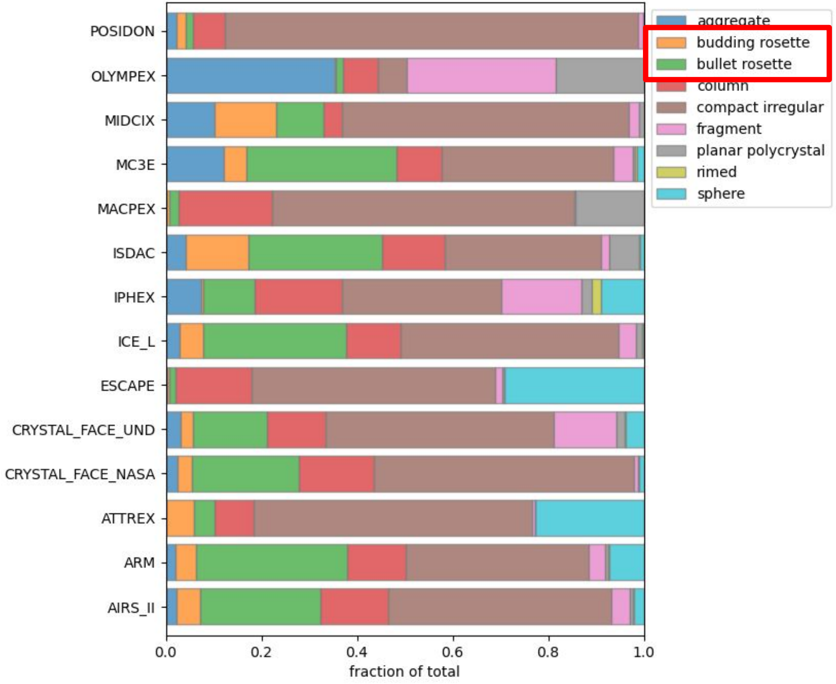
- Habit = Shape
- Habit ~ function of *temperature* and *supersaturation* (i.e., humidity)
- Habit influences:
    - **microphysical process rates**
    - **fall speeds**
    - **optical properties**
- E.g. Ice complexity may induce additional cooling effect of -1.1 W m^{-2} (Jarvinen et al. 2018)

Reconstructing 3-D crystals from CPI images

w/ Kara Sulia (U. Albany), Vanessa Przybylo (formerly U. Albany)

Methods



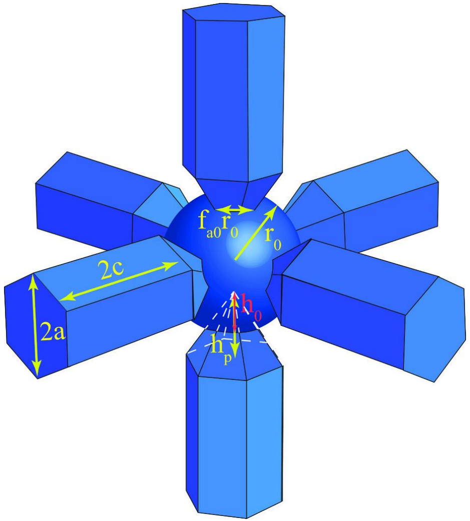


Credit: Przybylo et al. (2022)

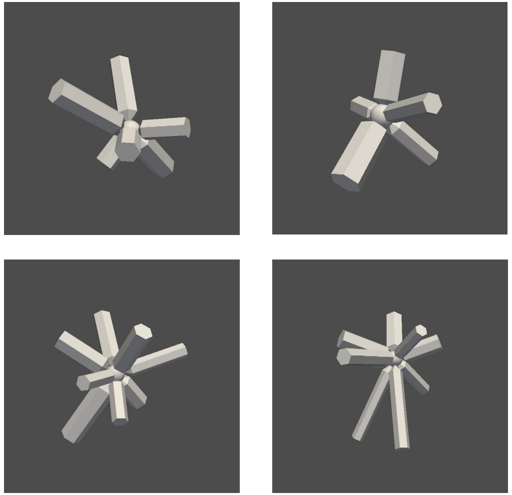
- CPI = Cloud Particle Imager
- Millions of CPI images from various aircraft field campaigns
- Lots of data, but limited to 2-D
- Ideally: 3-D features to constrain *mass-size relationships* → parameterizations
- **Basic idea: Train ML models that can extract 3-D features from 2-D images**

Synthetic 3-D dataset developed to train models

Source: Pokrifka et al., 2023



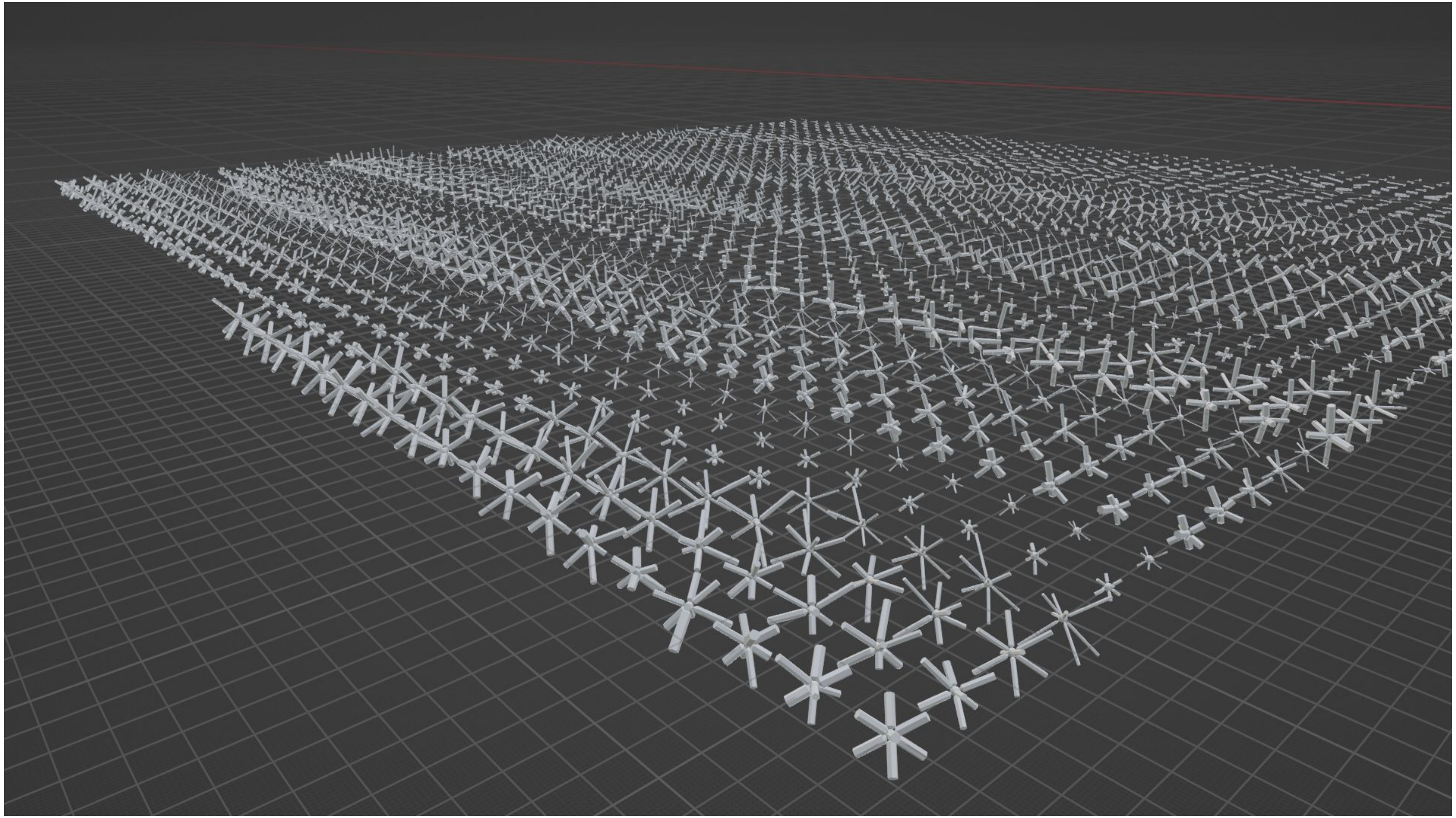
A priori
geometric model



Synthetic 3D models

- Bullet rosettes chosen as prototype; can be expanded to other habits
- Sphere size, arm aspect ratio, and angle of arms perturbed randomly
- Preliminary dataset of 9,000 crystals generated
- Developed in Python, code + dataset will be open-source and reproducible

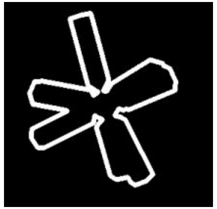
Synthetic 3-D dataset developed to train models



Overview of machine learning pipeline

Option 1: ML to predict 3D attributes

Extract 2d features from projections



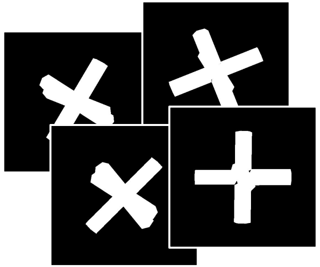
Data in tabular form

X (input): the 2d features

Y (output): e.g., effective density, surface area, # arms

Train models

Option 2: Deep learning for explicit reconstruction

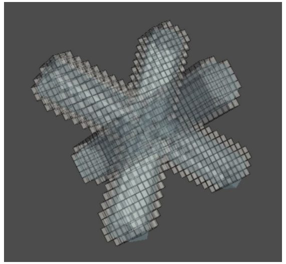


2-D projections from training set

Encoder (2D-NN)

Recurrence (3D-LSTM)

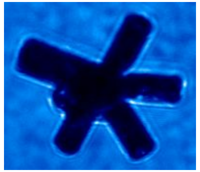
Decoder (3D-NN)



corresponding voxelized model

Test / Deploy

unseen data



Trained ML model

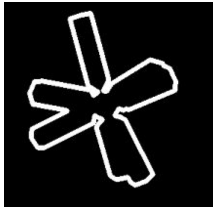
Predictions:
- mass
- # arms
- surface area
- etc.

Overview of machine learning pipeline

Methods

Option 1: ML to predict 3D attributes

Extract 2d features from projections



Data in tabular form

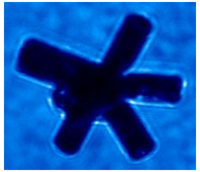
X (input): the 2d features

Y (output): e.g., effective density, surface area, # arms

Train models

Test / Deploy

unseen data



Trained ML model

Predictions:
- mass
- # arms
- surface area
- etc.

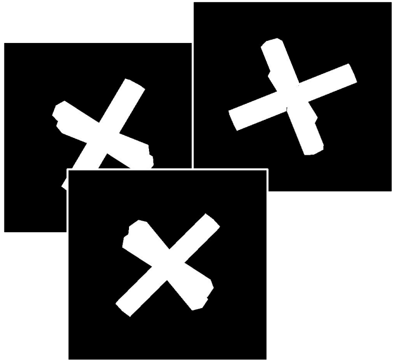
Details of Option 1: ML to predict 3D attributes

9,000
synthetic models

Calculate & save
target outputs:

(1) # arms,
(2) effective density
(3) effective surface area

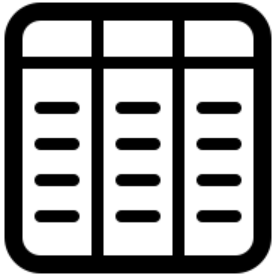
9,000 x 24 views
= 216,000 images (samples)



Take 24 random projections & calculate 2-D features for each projection:

(1) Aspect ratio
(2) Elliptical aspect ratio
(3) # extreme points
(4) Contour area
(5) Area ratio
(6) Complexity
(7) Circularity

Create merged
tabular dataset



- 216,000 rows
- 7 feature columns (inputs)
- 3 target columns (outputs)

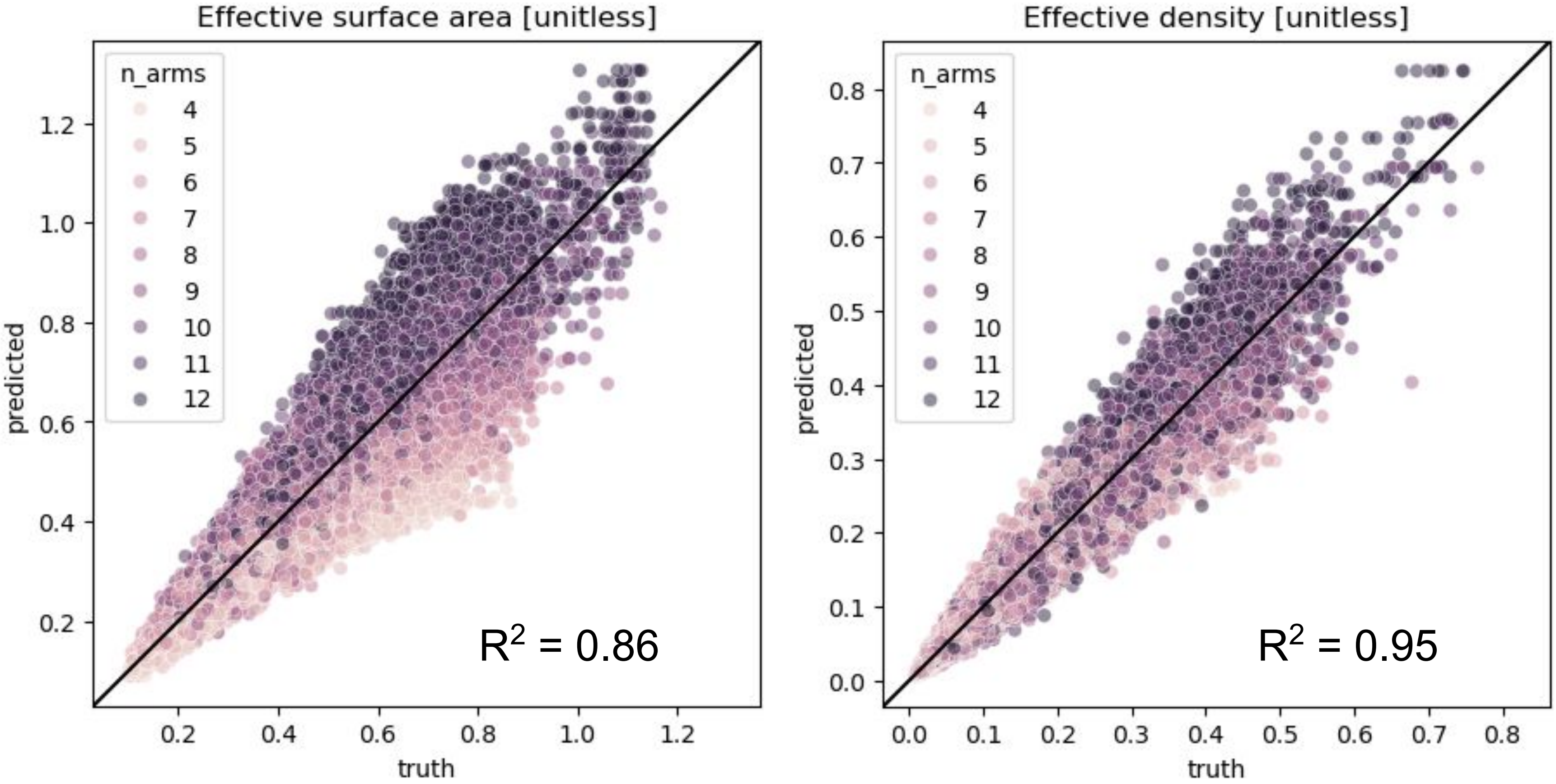
80%

Train model

20%

Test model

A random forest predicts effective surface area and density with moderate to high skill

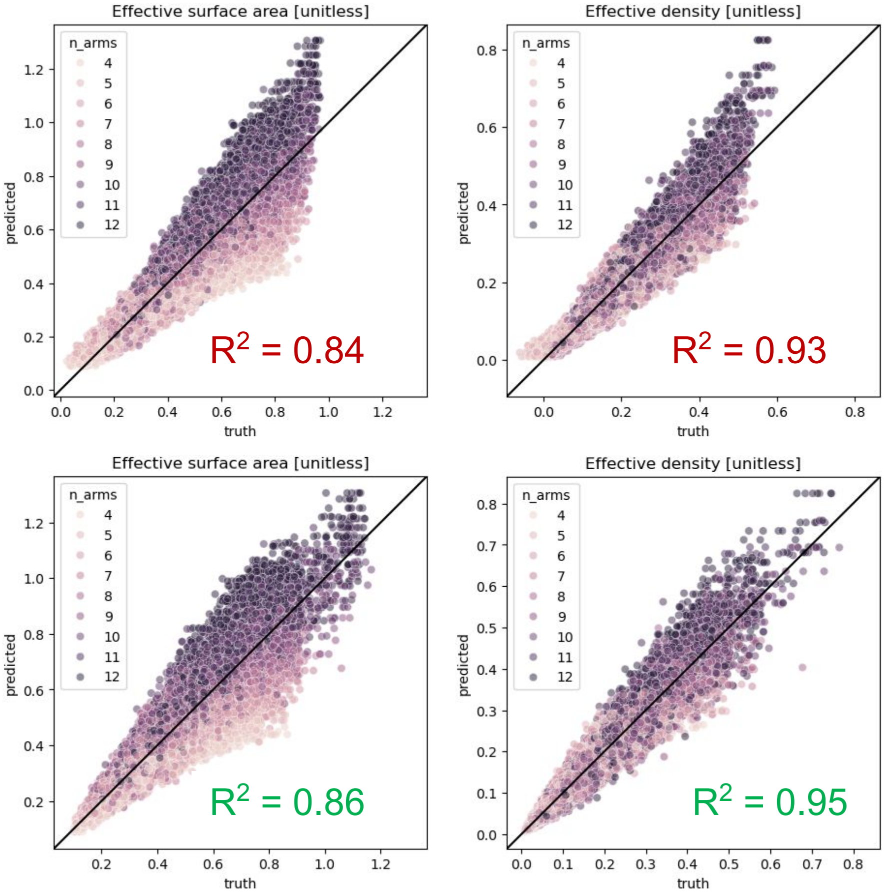


Random forest outperforms a linear regression

Results

Multivariate Linear Regression

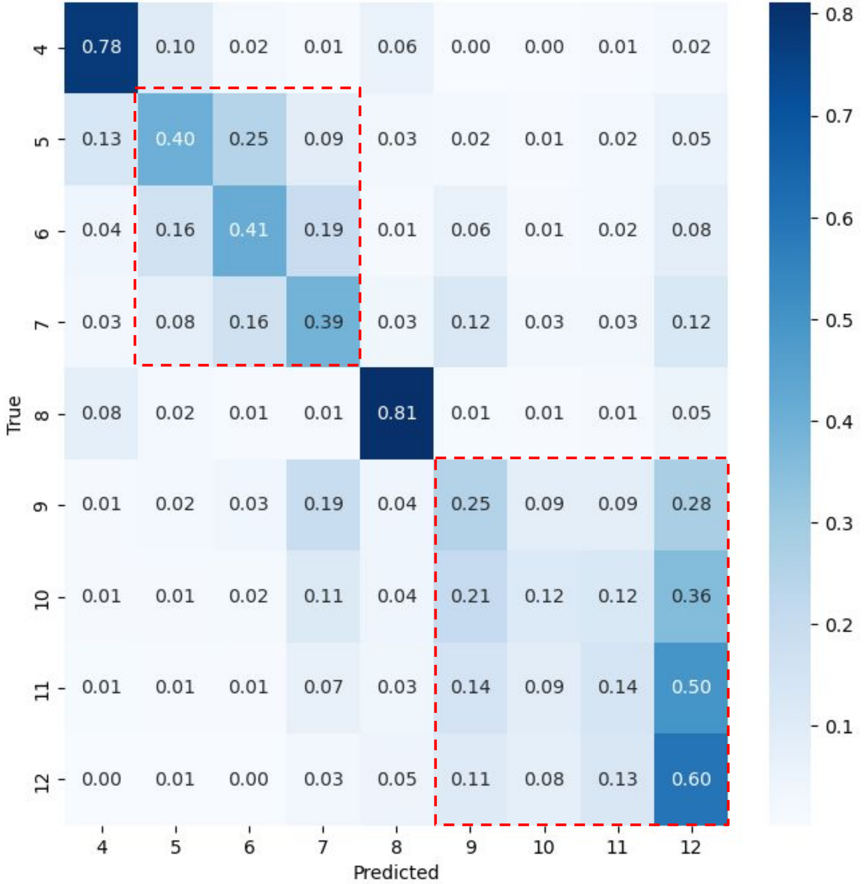
Random Forest



This indicates some non-linearities being captured by the random forest

RF predictions for # arms varies by class

Results

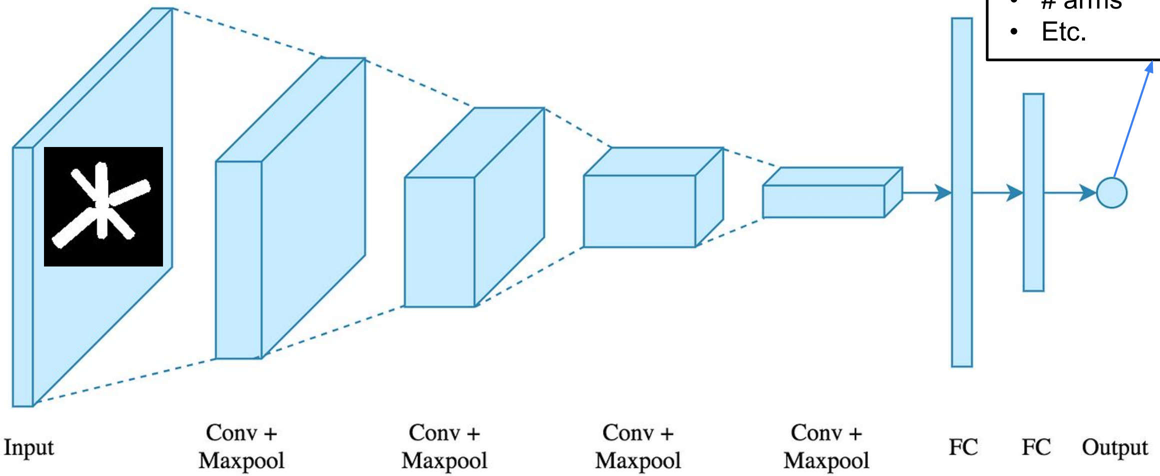


Confusion matrix for # of arms predicted by random forest

Normalized by row (i.e., each row sums up to 1.0)

A perfect predictor would show 1.0 values in the diagonal

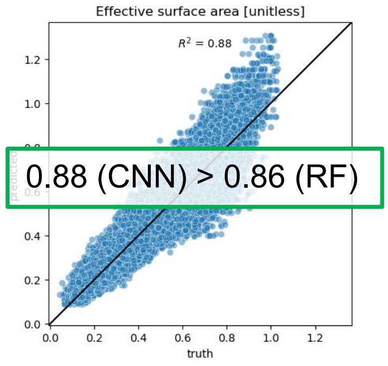
CNN can be used to predict 3-D targets directly from images without feature engineering

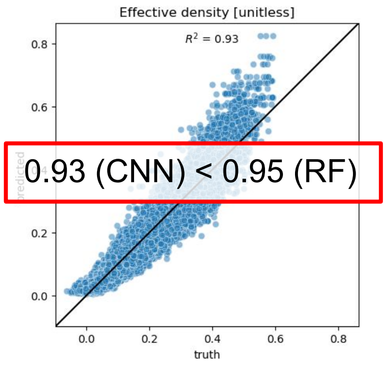


Credit: Arden Dertat

Predict:
- Effective density
- Effective surface area
- # arms
- Etc.

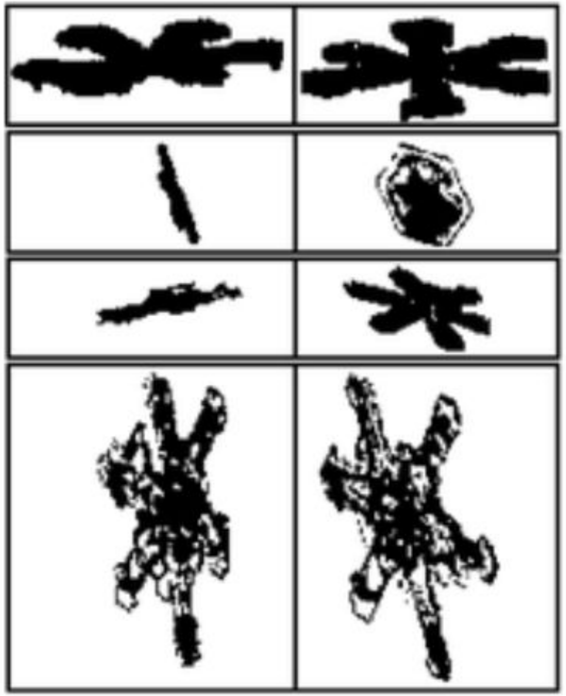
Preliminary results:





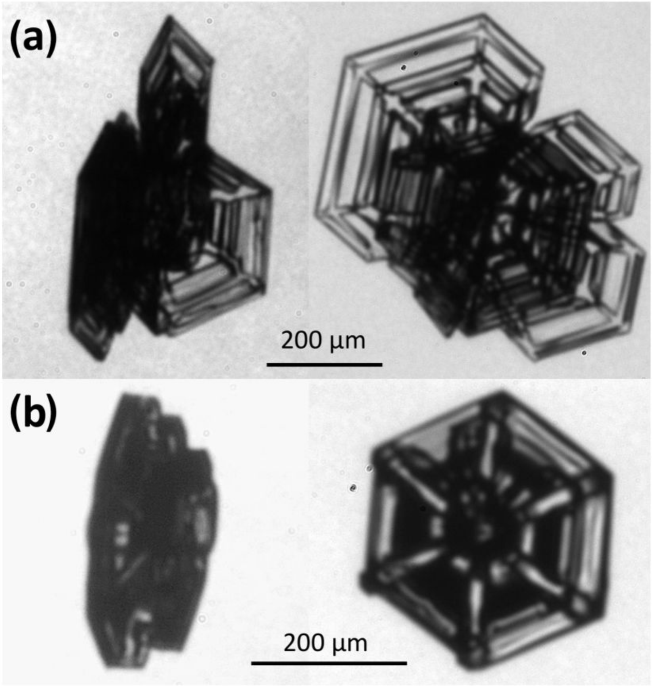
What if we add an additional view?

2D-S probe



Lawson et al. 2006

PHIPS-HALO



Schnaiter et al. 2018

Pipeline w/ two views

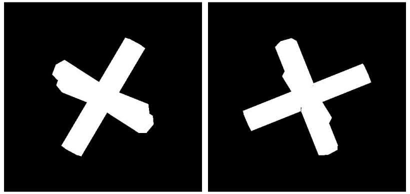
9,000
synthetic models

Calculate & save target outputs:

(1) # arms,
(2) effective density
(3) effective surface area

9,000 x ***24 image pairs***
= 216,000 image pairs

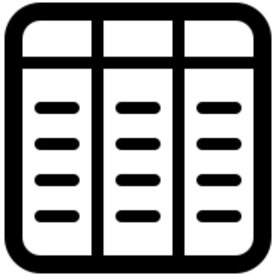
View 1 View 2



Take 24 random ***projection pairs*** & calculate 2-D features for each projection image of each pair:

(1) Aspect ratio
(2) Elliptical aspect ratio
(3) # extreme points
(4) Contour area
(5) Area ratio
(6) Complexity
(7) Circularity

Create merged tabular dataset



- 216,000 rows
- ***7x2 = 14 feature columns (inputs)***
- 3 target columns (outputs)

80%

Train model

20%

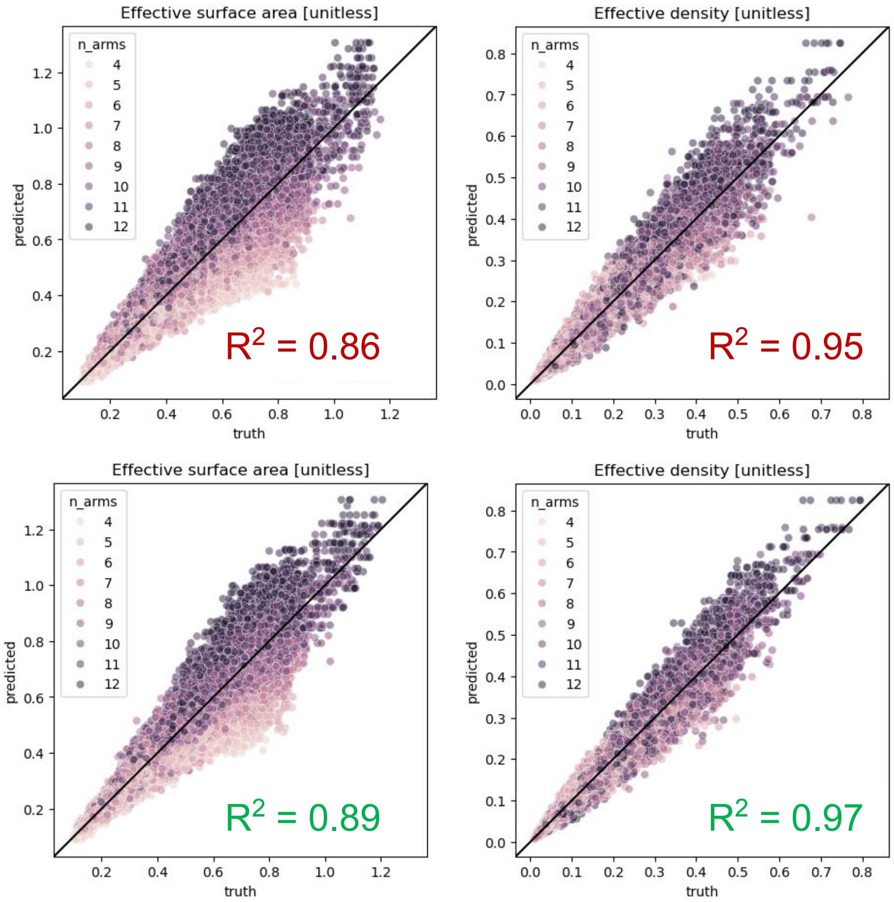
Test model

Two views are better than one

Results

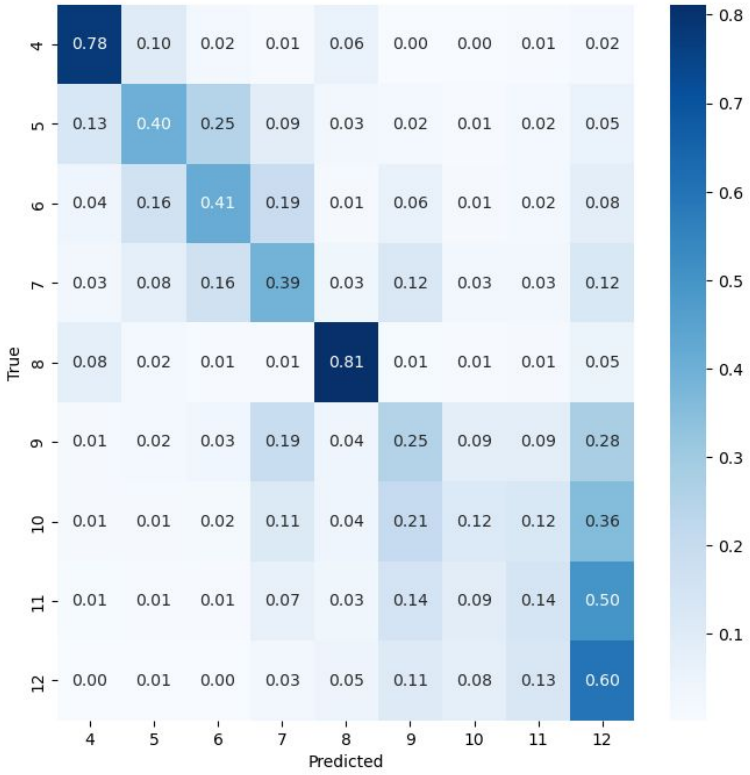
Random Forest
w/ single view

Random Forest
w/ two views

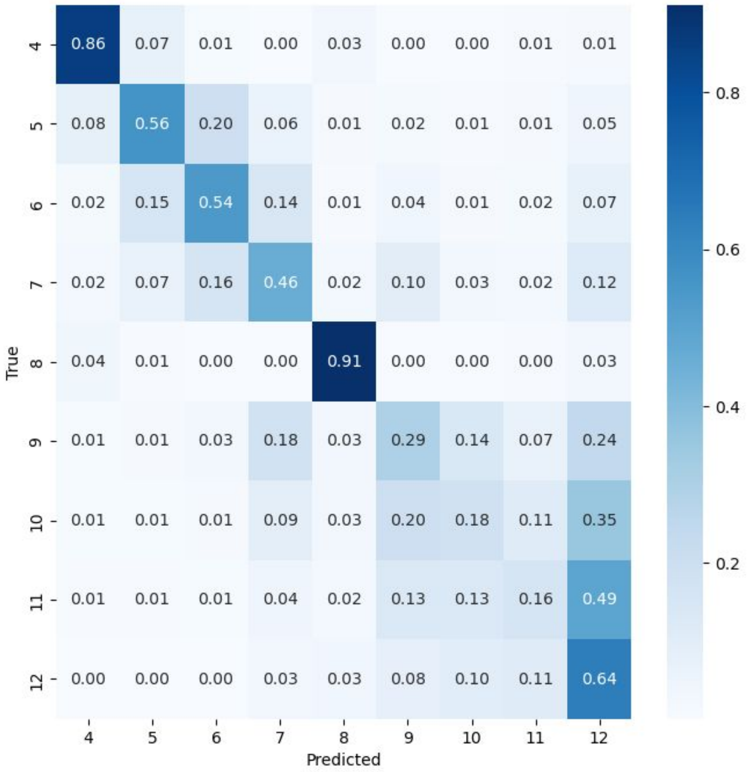


Two views are better than one

Single view



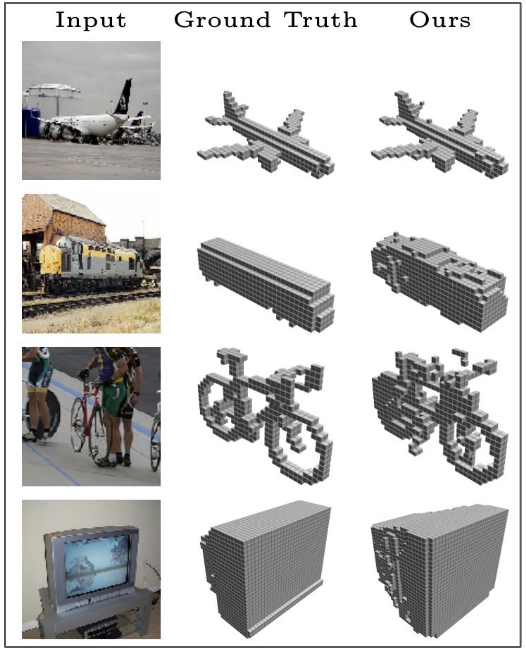
Two views



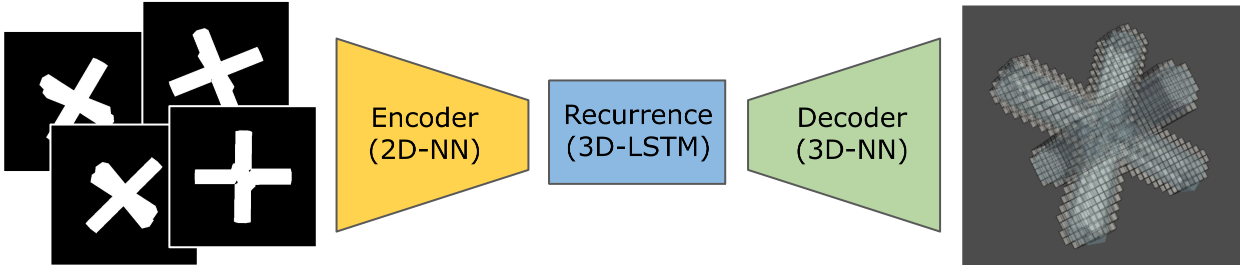
Deep learning: explicit reconstruction

Ongoing work

3D Recurrent Reconstruction Neural Network (3D-R2N2):
Choy et al., 2016



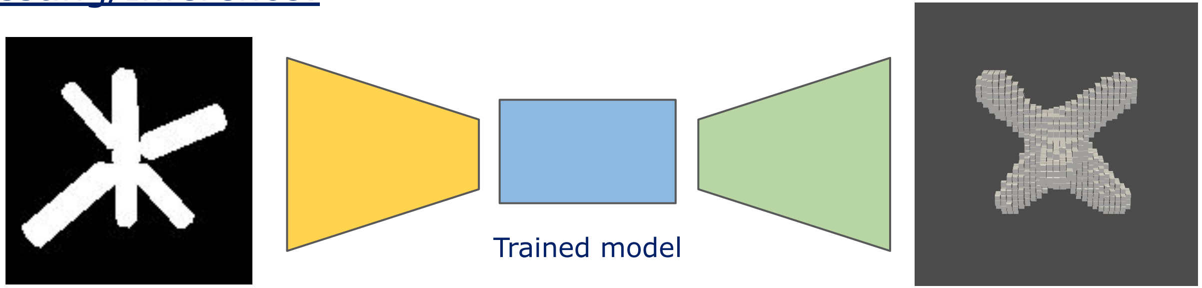
Training:



2-D projections from training set

corresponding voxelized model

Testing/Inference:

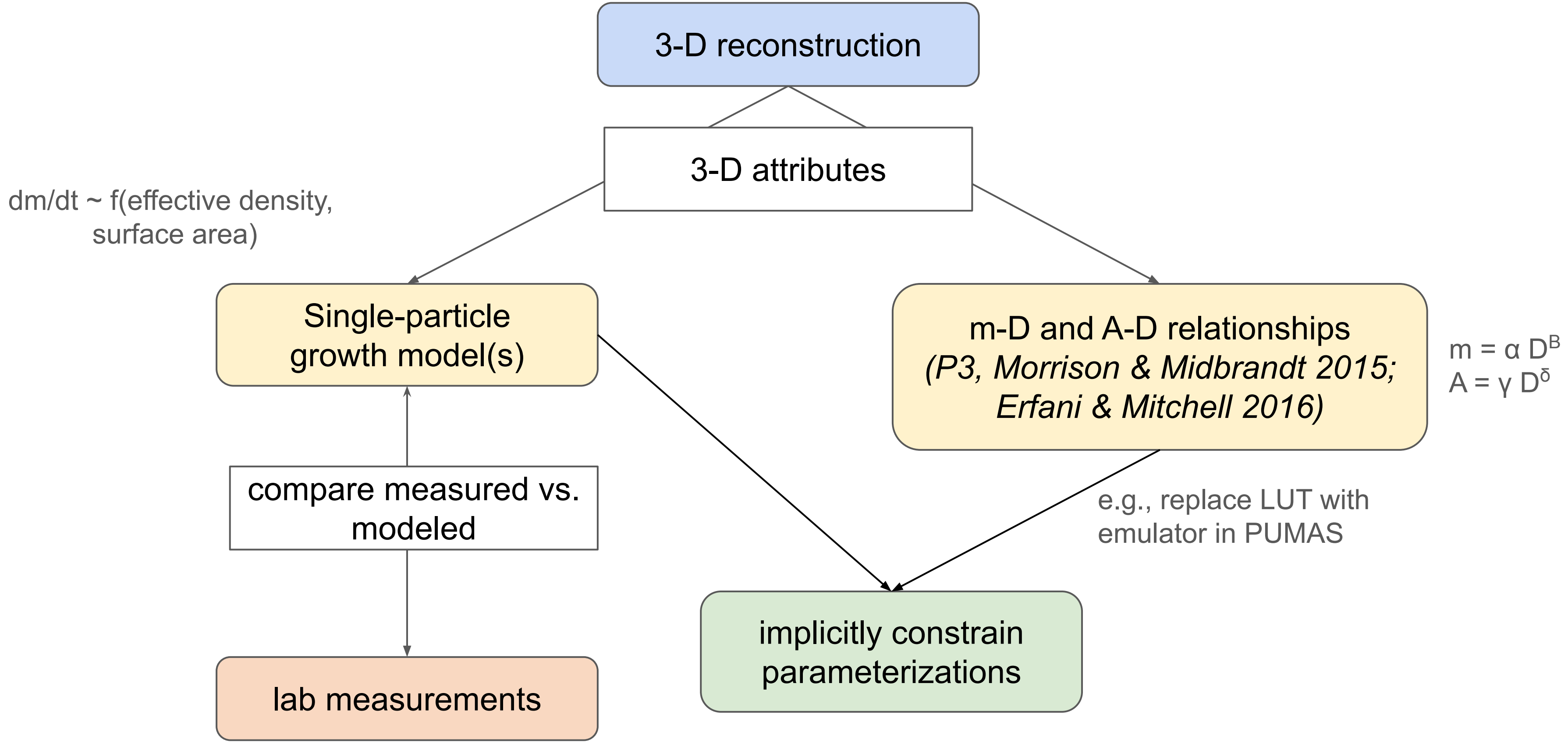


Testing: unseen projections
Inference: CPI images

Trained model

inferred voxelized model

Constraining 3-D microphysics properties □ better parameterizations



Conclusion

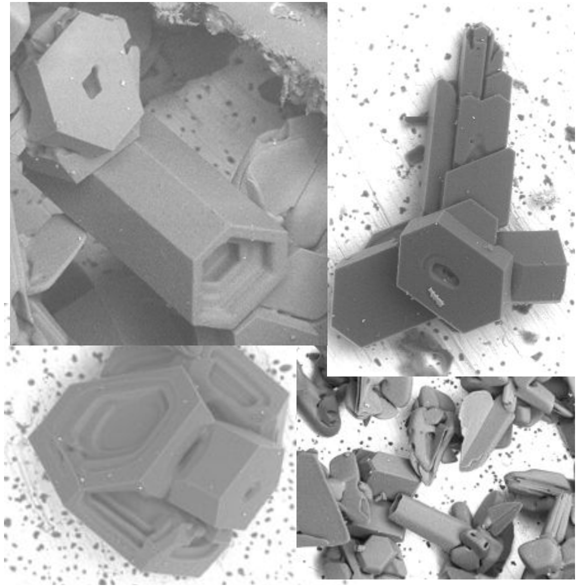
1. A dataset of synthetic bullet rosette meshes was created
2. ML was able to predict effective density and surface area with encouraging skill (to be improved)
3. The classification of # arms was more challenging
4. Inferring 3-D properties from CPI images will allow us to improve parameterizations moving forward

Thanks!

Contact: jk4730@columbia.edu

Advisors, collaborators & help:

Kara Lamb (Columbia)
Marcus van Lier-Walqui (Columbia, NASA GISS)
Jerry Harrington (Penn State)
Kamal Chandrakar (NCAR)
Hugh Morrison (NCAR)
Jonas Mikhaeli (Columbia)
Kaitlyn Loftus (Columbia)
Kara Sulia (U. Albany)
Vanessa Przybylo (formerly U. Albany)
…et al.



Images from Nathan Magee

Funding: LEAP NSF U.S. DEPARTMENT OF ENERGY

Conclusion

Contact: jk4730@columbia.edu

1. A dataset of synthetic bullet rosette meshes was created
2. ML was able to predict effective density and surface area with encouraging skills (to be improved)
3. The classification of # arms was more challenging
4. Inferring 3-D properties from CPI images will allow us to improve parametrizations moving forward