

# Convolutions on Spherical Images

Marc Eder and Jan-Michael Frahm

*Department of Computer Science, University of North Carolina at Chapel Hill*

*2019 SUMO Workshop: 360° Indoor Scene Understanding and Modeling*

*CVPR 2019, Long Beach, CA*

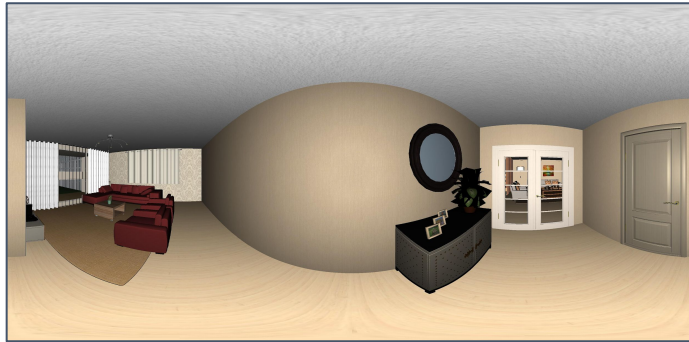
*June 17, 2019*



UNC Computer Vision

THE UNIVERSITY  
of NORTH CAROLINA  
at CHAPEL HILL





# Image representation matters!

Simply resampling the image to a different representation significantly improves accuracy for predictions tasks with convolutional neural networks.

# Why does image representation matter?

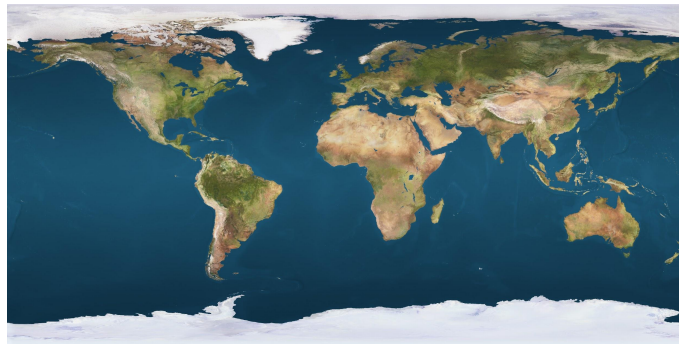
Gauss's Theorema Egregium:

*Gaussian curvature of a surface is invariant under local isometry*

Far reaching implications, but particularly relevant to cartography: **All planar projections of a sphere have distortions**



*Spherical Earth Model*



*A Distorted Map Projection*



*Carl Friedrich Gauss*

# All 360° image representations are distorted

## Cubemap

### *Gnomonic (rectilinear) projection*

- Popular graphics format
- Projects a sphere onto the faces of an inscribing cube
- Distorts most severely in corners of faces



## Equirectangular image

### *Equirectangular projection*

- Simple transformation from sphere to projection
- Indexes image grid with spherical coordinates
- Distorts most severely near poles



# So what?

Why do we care about spherical distortion when using CNNs?

# Distortion and convolution

## 1D Discrete Convolution

$$(f * g)[n] = \sum_{m=-\lfloor \frac{K}{2} \rfloor}^{\lfloor \frac{K}{2} \rfloor} f[m]g[n - m]$$

Separating the sampling operation  
from the weighted summation

$$= \sum_{m=-\lfloor \frac{K}{2} \rfloor}^{\lfloor \frac{K}{2} \rfloor} f[m] \left( \sum_{l=-\infty}^{\infty} g[l] \delta[l - n + m] \right)$$

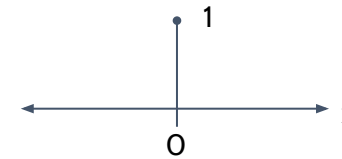


# Distortion and convolution

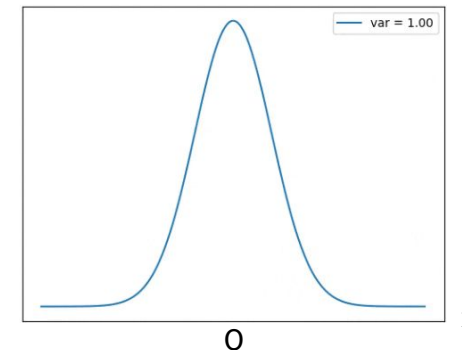
$$(f * g)[n] = \sum_{m=-\lfloor \frac{K}{2} \rfloor}^{\lfloor \frac{K}{2} \rfloor} f[m] \left( \sum_{l=-\infty}^{\infty} g[l] \delta[l - n + m] \right)$$

Sampling represented by the Dirac delta function

**Dirac delta function:**  $\delta[x] = \begin{cases} 1 & x = 0 \\ 0 & o.w. \end{cases}$



**Alternatively:**  $\delta(x) = \lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$   
(in continuous form)



(area = 1)

# Distortion and convolution

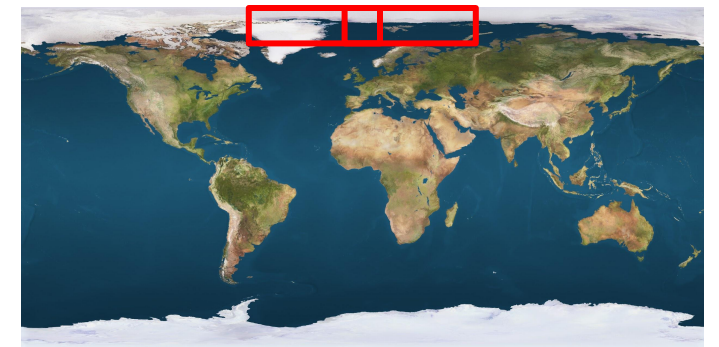
$$(f * g)[n] = \sum_{m=-\lfloor \frac{K}{2} \rfloor}^{\lfloor \frac{K}{2} \rfloor} f[m] \left( \sum_{l=-\infty}^{\infty} g[l] \delta[l - n + m] \right)$$

← Adds unexpected scaling bias

**Key observation:** Translational equivariance implicitly assumes all sampled data contribute equal information

***Spherical distortion violates this assumption***

*E.g. Pixel redundancy at poles in equirectangular image*






# How can we fix this?

Let's look at what cartographers do...

# The imperfect map


WHAT YOUR FAVORITE  
**MAP PROJECTION**  
SAYS ABOUT YOU

**MERCATOR**




YOU'RE NOT REALLY INTO MAPS.

**ROBINSON**




YOU HAVE A COMFORTABLE PAIR OF RUNNING SHOES THAT YOU WEAR EVERYWHERE. YOU LIKE COFFEE AND ENJOY THE BEATLES. YOU THINK THE ROBINSON IS THE BEST-LOOKING PROJECTION, HANDS DOWN.

**VAN DER GRINTEN**




YOU'RE NOT A COMPLICATED PERSON. YOU LOVE THE MERCATOR PROJECTION; YOU JUST WISH IT WEREN'T SQUARE. THE EARTH'S NOT A SQUARE, IT'S A CIRCLE. YOU LIKE CIRCLES. TODAY IS GONNA BE A GOOD DAY!

**DYMAXION**




YOU LIKE ISAAC ASIMOV, XML, AND SHOES WITH TOES. YOU THINK THE SEGWAY GOT A BAD RAP. YOU OWN 3D GOGGLES, WHICH YOU USE TO VIEW ROTATING MODELS OF BETTER 3D GOGGLES. YOU TYPE IN DVORAK.

**WINKEL-TRIPEL**




NATIONAL GEOGRAPHIC ADOPTED THE WINKEL-TRIPEL IN 1998, BUT YOU'VE BEEN A WIKI FAN SINCE LONG BEFORE "NAT GED" SHOWED UP. YOU'RE WORRIED IT'S GETTING PLAYED OUT, AND ARE THINKING OF SWITCHING TO THE KAVRAYSKIY. YOU ONCE LEFT A PARTY IN DISGUST WHEN A GUEST SHOWED UP WEARING SHOES WITH TOES. YOUR FAVORITE MUSICAL GENRE IS "POST-".

**GOODE HOMOLOSINE**



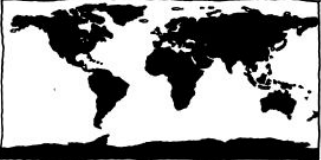
THEY SAY MAPPING THE EARTH ON A 2D SURFACE IS LIKE FLATTENING AN ORANGE PEEL, WHICH SEEMS EASY ENOUGH TO YOU. YOU LIKE EASY SOLUTIONS. YOU THINK WE WOULDN'T HAVE SO MANY PROBLEMS IF WE'D JUST ELECT *NORMAL* PEOPLE TO CONGRESS INSTEAD OF POLITICIANS. YOU THINK AIRLINES SHOULD JUST BUY FOOD FROM THE RESTAURANTS NEAR THE GATES AND SERVE *THAT* ON BOARD. YOU CHANGE YOUR CAR'S OIL, BUT SECRETLY WONDER IF YOU REALLY *NEED* TO.

**HOBBO-DYER**




YOU WANT TO AVOID CULTURAL IMPERIALISM, BUT YOU'VE HEARD BAD THINGS ABOUT GALL-PETERS. YOU'RE CONFLICT-AVERSE AND BUY ORGANIC. YOU USE A RECENTLY-INVENTED SET OF GENDER-NEUTRAL PRONOUNS AND THINK THAT WHAT THE WORLD NEEDS IS A REVOLUTION IN CONSCIOUSNESS.

**PLATE CARRÉE (EQUIRECTANGULAR)**




YOU THINK THIS ONE IS FINE. YOU LIKE HOW X AND Y MAP TO LATITUDE AND LONGITUDE. THE OTHER PROJECTIONS OVERCOMPLICATE THINGS. YOU WANT ME TO STOP ASKING ABOUT MAPS SO YOU CAN ENJOY DINNER.

**A GLOBE!**




YES, YOU'RE VERY CLEVER.

**WATERMAN BUTTERFLY**



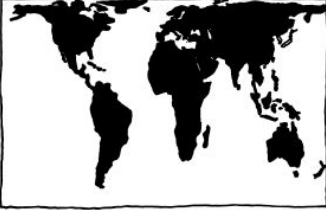
REALLY? YOU KNOW THE WATERMAN? HAVE YOU SEEN THE 1909 CAHILL MAP IT'S BASED — ... YOU HAVE A FRAMED REPRODUCTION AT HOME?! WHOA. ... LISTEN, FORGET THESE QUESTIONS. ARE YOU DOING ANYTHING TONIGHT?

**PEIRCE QUINCUNCIAL**



YOU THINK THAT WHEN WE LOOK AT A MAP, WHAT WE REALLY SEE IS OURSELVES. AFTER YOU FIRST SAW *INCEPTION*, YOU SAT SILENT IN THE THEATER FOR SIX HOURS. IT BREAKS YOU OUT TO REALIZE THAT EVERYONE AROUND YOU HAS A SKELETON INSIDE THEM. YOU *HAVE* REALLY LOOKED AT YOUR HANDS.

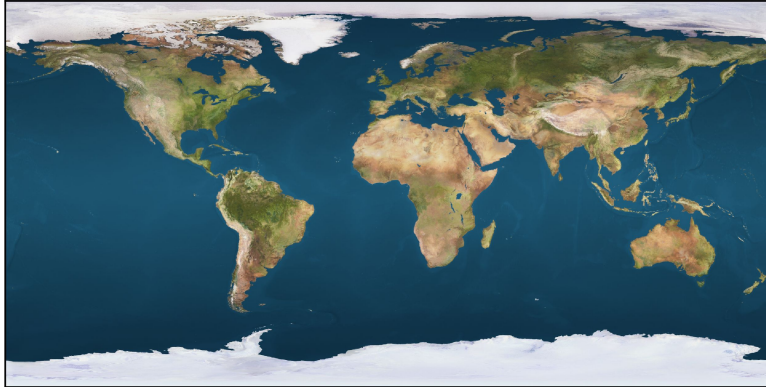
**GALL-PETERS**



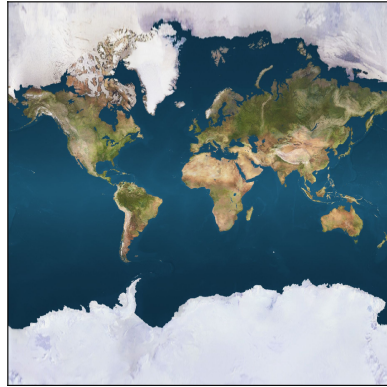
I HATE YOU.

Cropped from <https://xkcd.com/977/>

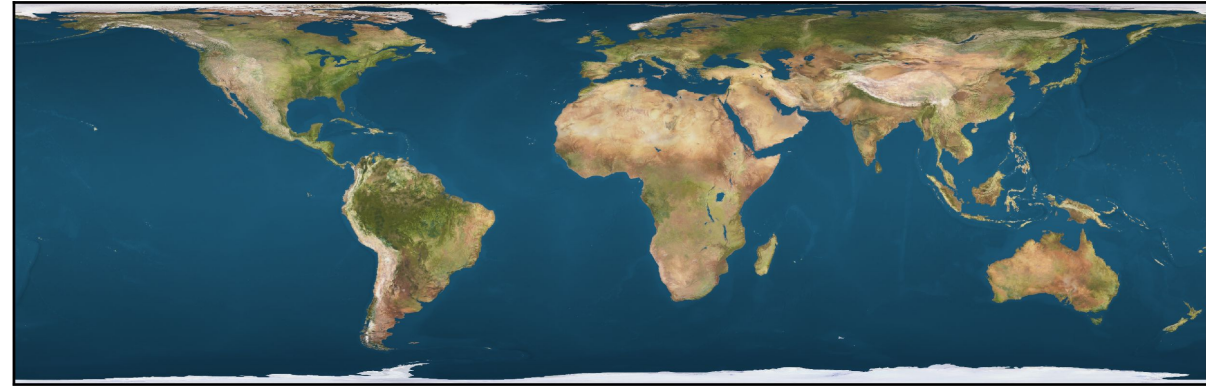
# Analyzing spherical distortion



**Equidistant**  
Preserves distances between points  
(*Equirectangular*)

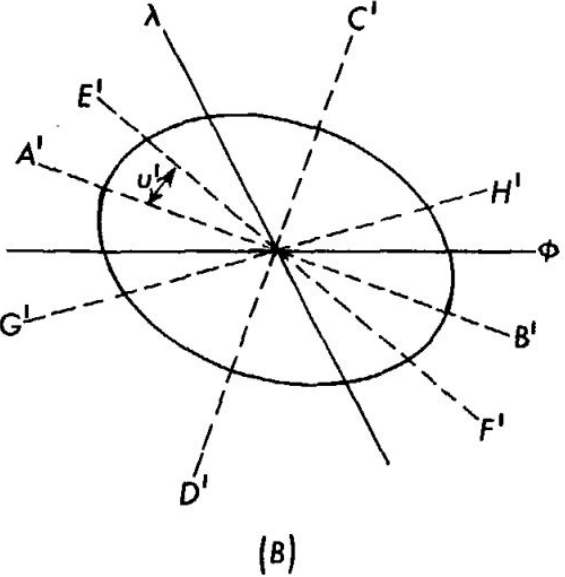
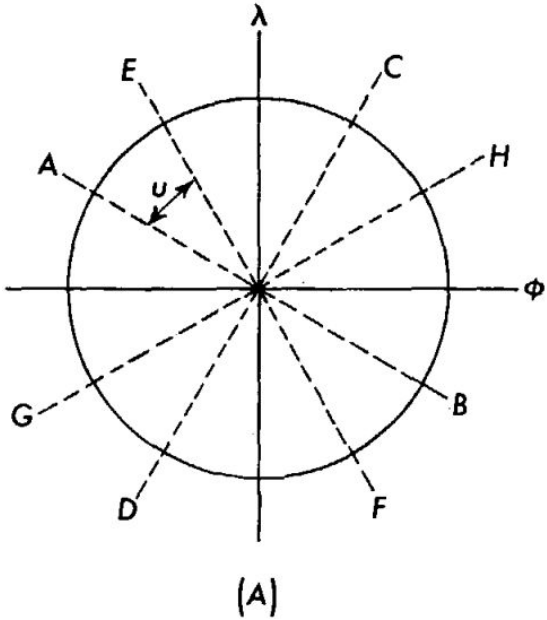


**Conformal**  
Preserves local angles  
(*Mercator*)



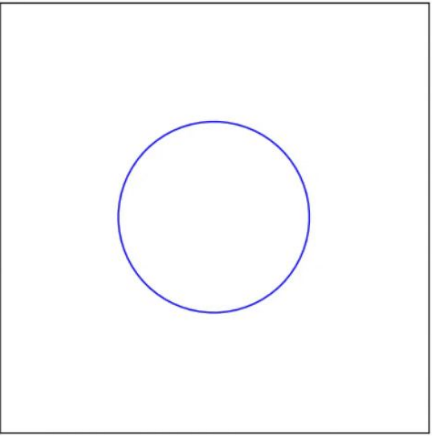
**Equal Area**  
Preserves relative sizes of objects  
(*Gall-Peters*)

# Analyzing spherical distortion



**Tissot's Indicatrix:** An infinitely small circle on the Earth (A) appears as an ellipse on a typical map (B)

*Recall modeling convolution's sampling function as the limit of a Gaussian as  $\sigma \rightarrow 0$*

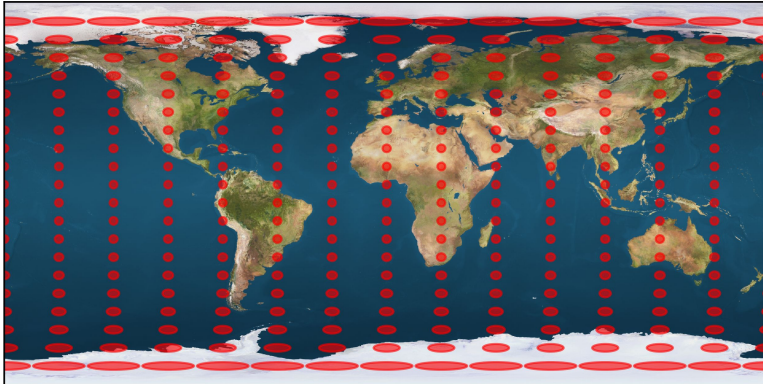


*2D Gaussian as  $\sigma \rightarrow 0$*

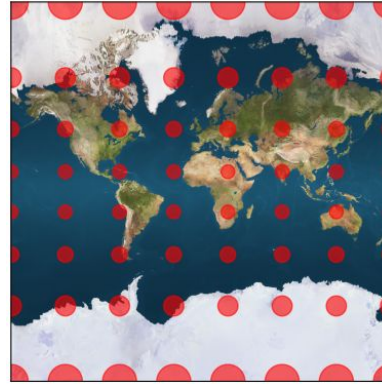
Tissot figure from Snyder, John Parr. *Map projections--A working manual*. Vol. 1395. US Government Printing Office, 1987.



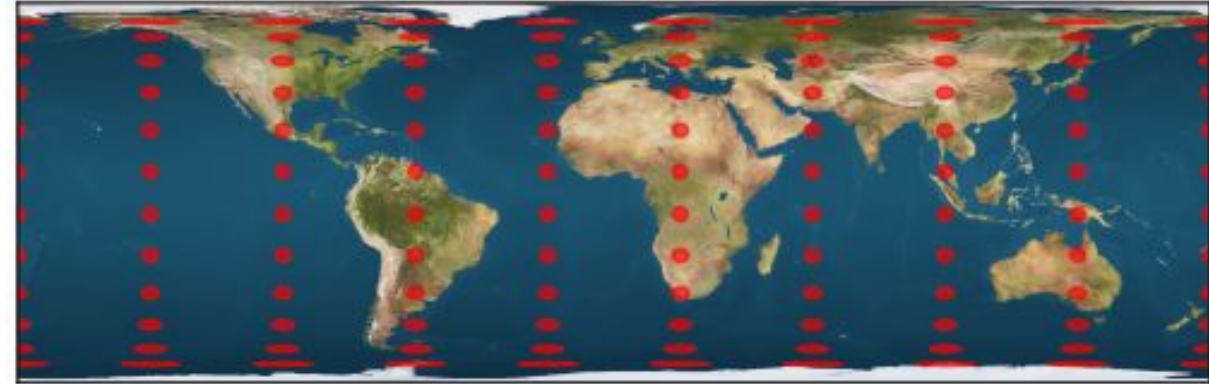
# Analyzing spherical distortion



Equidistant  
Preserves distances between points  
(*Equirectangular*)



Conformal  
Preserves local angles  
(*Mercator*)



Equal Area  
Preserves relative sizes of objects  
(*Gall-Peters*)

# Back to spherical images

Let's take another look at those two common spherical image formats...

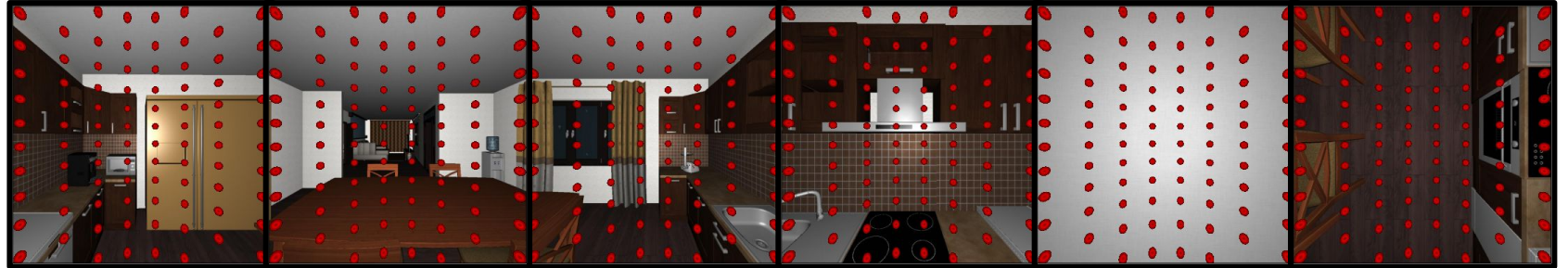


# Distortion in 360° image representations

## Cubemap

### *Gnomonic (rectilinear) projection*

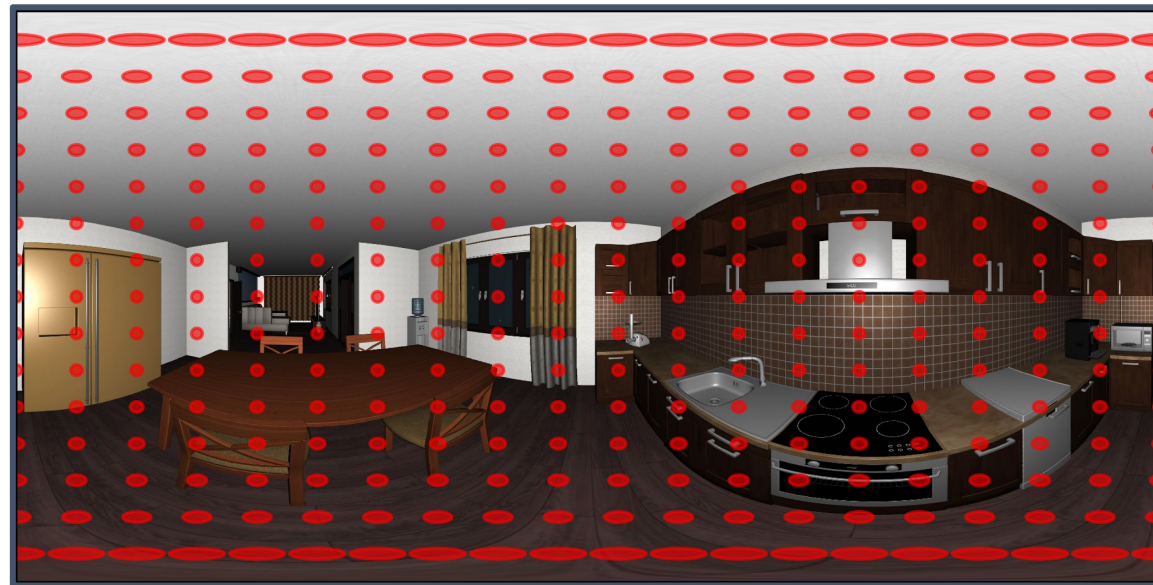
- Popular graphics format
- Projects a sphere onto the faces of an inscribing cube
- **Distorts most severely in corners of faces**



## Equirectangular image

### *Equirectangular projection*

- Simple transformation from sphere to projection
- Indexes image grid with spherical coordinates
- **Distorts most severely near poles**

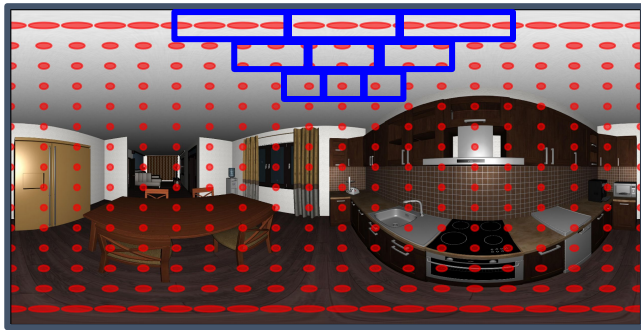


# Quick summary of spherical distortion

1. Mathematically impossible to remove
2. Disrupts *translational equivariance* critical to CNN function
3. Spreads and deforms content (information) in images

# Two solutions

## Accumulate deformed content



*Example accumulation kernel*

### Pros:

- Works with any image representation

### Cons:

- Very inefficient (possibly >100's of pixels per sample)
- GPU implementation difficult

## Use a compromise projection



*Planar approximation to sphere*

### Pros:

- Efficient sampling (just a single pixel)
- Can use standard grid convolution with limited modifications to implementation

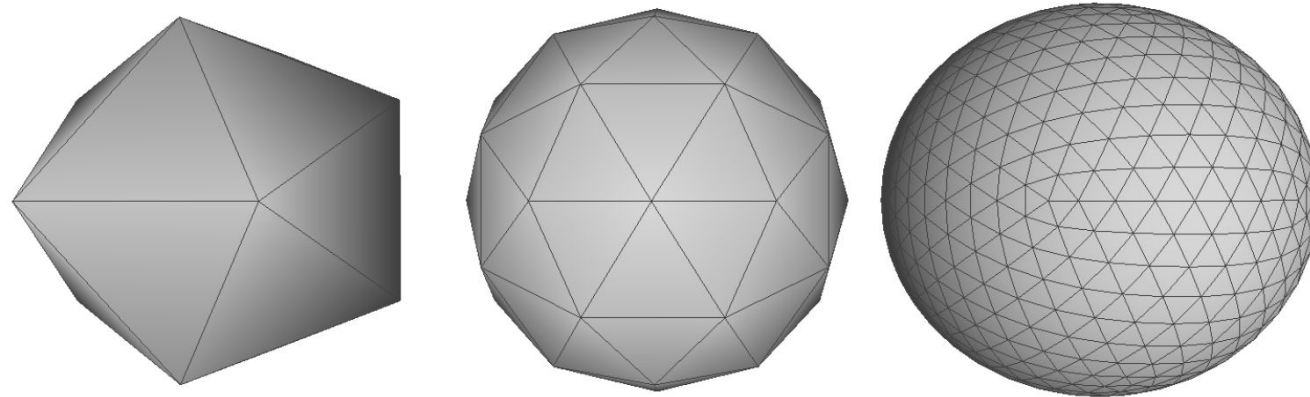
### Cons:

- Some distortion remains

# ISEA and the icosphere

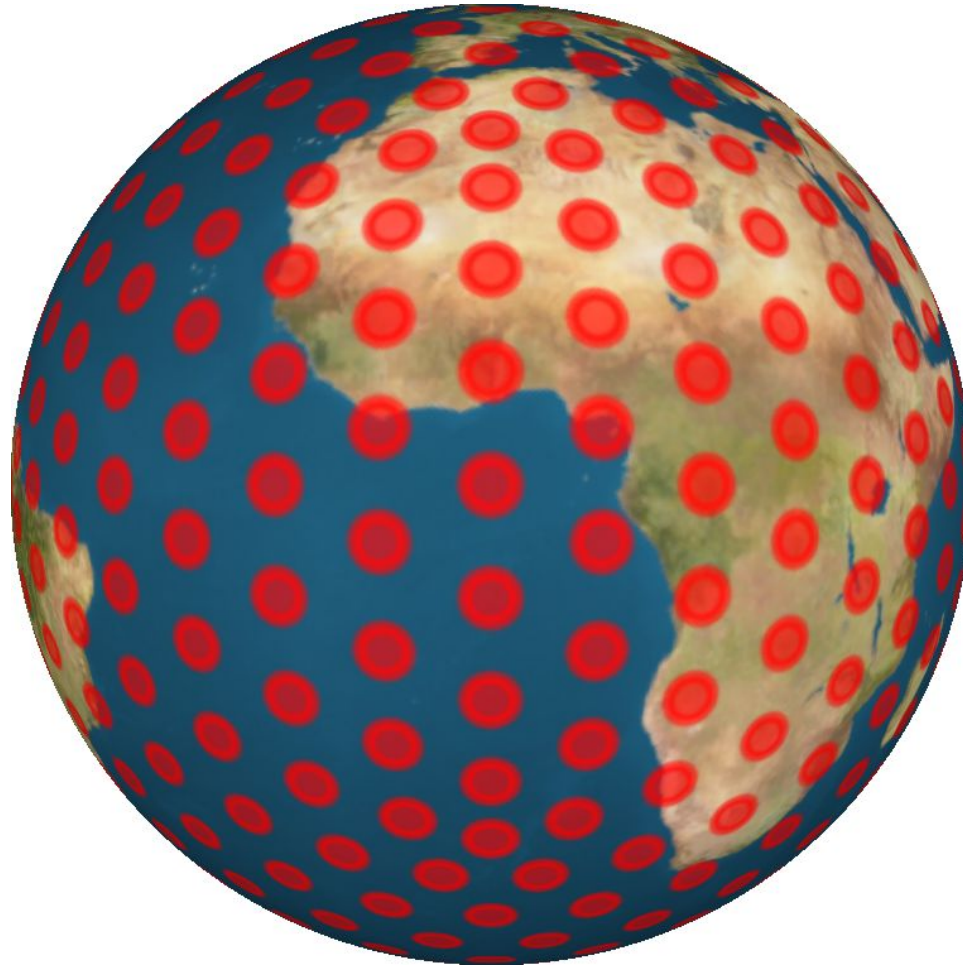
Our compromise projection: Icosahedral Snyder equal area (ISEA) projection [3]

Projects image onto surface of **icosphere**, a recursively subdivided regular icosahedron



One of least distorted compromise projections [2]

# ISEA and the icosphere



# Evaluation

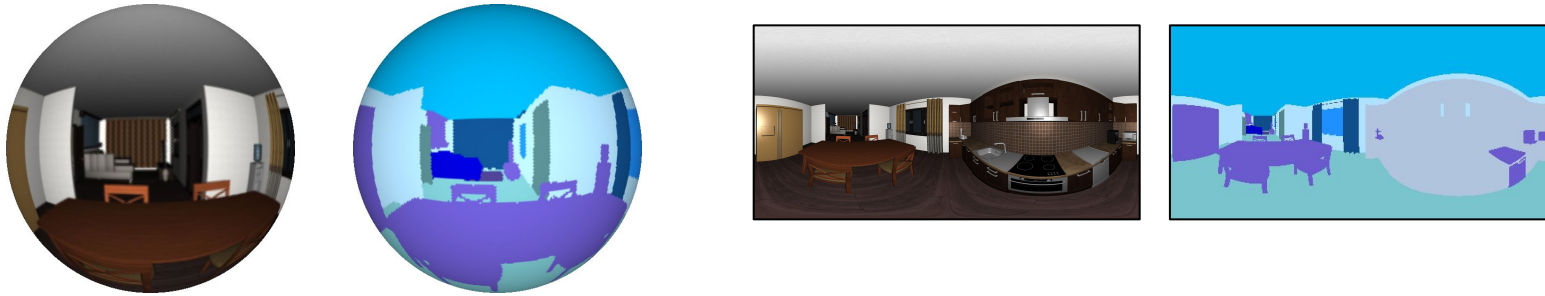
Semantic segmentation improves 12.6% simply due to change of image representation



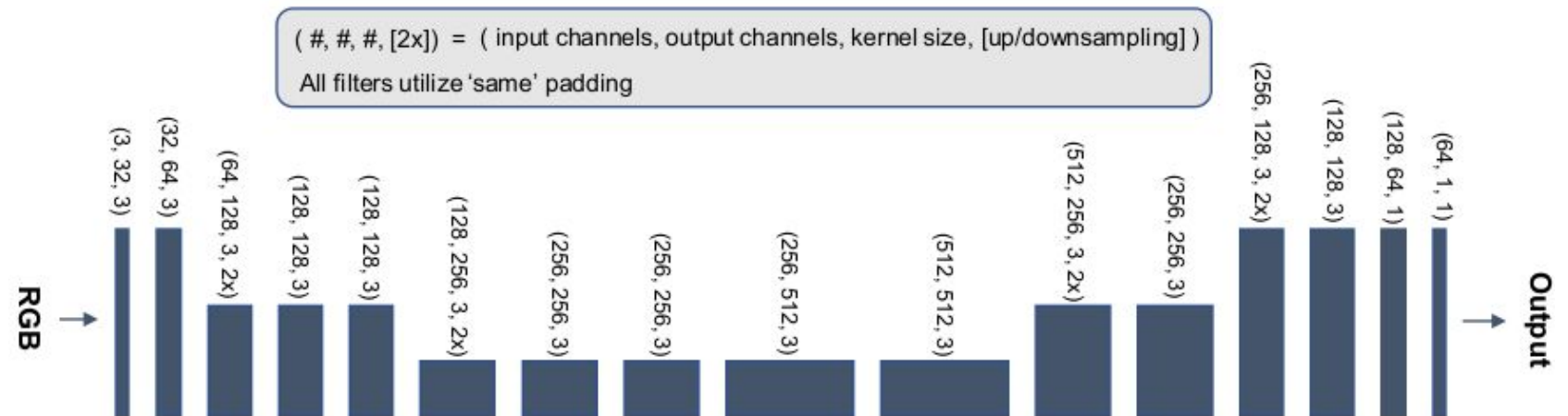


# Semantic segmentation

Train a network with each representation using SUMO dataset [5]



Simple encoder-decoder



# Results

Evaluate mIOU on 15 most frequent semantic classes

Representation	Floor	Ceiling	Wall	Door	Cabinet	Rug	Window	Curtain
<i>Equirectangular (Gnom. Kernel) [1, 3]</i>	0.9315	<b>0.9710</b>	0.8597	0.6466	0.6376	<b>0.7284</b>	0.7012	0.4703
<i>Icosphere (ours)</i>	<b>0.9352</b>	0.9703	<b>0.8797</b>	<b>0.6890</b>	<b>0.7037</b>	0.6970	<b>0.7562</b>	<b>0.5744</b>

Representation	Sofa	Partition	Bed	Chair	Table	Shelving	Chandelier	All Classes
<i>Equirectangular (Gnom. Kernel) [1, 3]</i>	0.7114	0.4172	0.7133	0.4219	0.4587	0.3278	<b>0.4491</b>	0.5904
<i>Icosphere (ours)</i>	<b>0.7374</b>	<b>0.4683</b>	<b>0.7776</b>	<b>0.4375</b>	<b>0.5018</b>	<b>0.3733</b>	0.4472	<b>0.6639</b>

**ISEA projection gives a 12.6% improvement over state-of-the-art methods that use equirectangular images!**

# Other applications and future work

## Not limited to CNNs

Normalized correlation metrics suffer from same issues with spherical images (e.g. stereo depth)

Image filtering uses convolution too -- 360° panos are a growing social media commodity (e.g. Instagram filters)

Need to build large-scale *realistic* spherical image dataset

# Thank you!

## Any questions?

For more conversation, come to our poster today or contact Marc Eder at [meder@cs.unc.edu](mailto:meder@cs.unc.edu).

# References

## Images:

- Equirectangular Earth image, used with permission from <http://planetpixelemporium.com/earth8081.html>
- Gauss, slide 3, from [https://en.wikipedia.org/wiki/Carl\\_Friedrich\\_Gauss#/media/File:Carl\\_Friedrich\\_Gauss\\_1840\\_by\\_Jensen.jpg](https://en.wikipedia.org/wiki/Carl_Friedrich_Gauss#/media/File:Carl_Friedrich_Gauss_1840_by_Jensen.jpg) (public domain)
- Map projection comic, slide 10, from <https://xkcd.com/977/> (creative commons license)
- Tissot indicatrix, slide 12, from Snyder, John Parr. Map projections--A working manual. Vol. 1395. US Government Printing Office, 1987.
- SUMO dataset images [5]

## Citations:

- [1] Coors, Benjamin, Alexandru Paul Condurache, and Andreas Geiger. "Spherenet: Learning spherical representations for detection and classification in omnidirectional images." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
- [2] Kimerling, Jon A., et al. "Comparing geometrical properties of global grids." Cartography and Geographic Information Science 26.4 (1999): 271-288.
- [3] Snyder, John P. "An equal-area map projection for polyhedral globes." Cartographica: The International Journal for Geographic Information and Geovisualization 29.1 (1992): 10-21.
- [4] Tateno, Keisuke, Nassir Navab, and Federico Tombari. "Distortion-aware convolutional filters for dense prediction in panoramic images." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
- [5] Tchapmi, Lyne and Daniel Huber. The sumo challenge