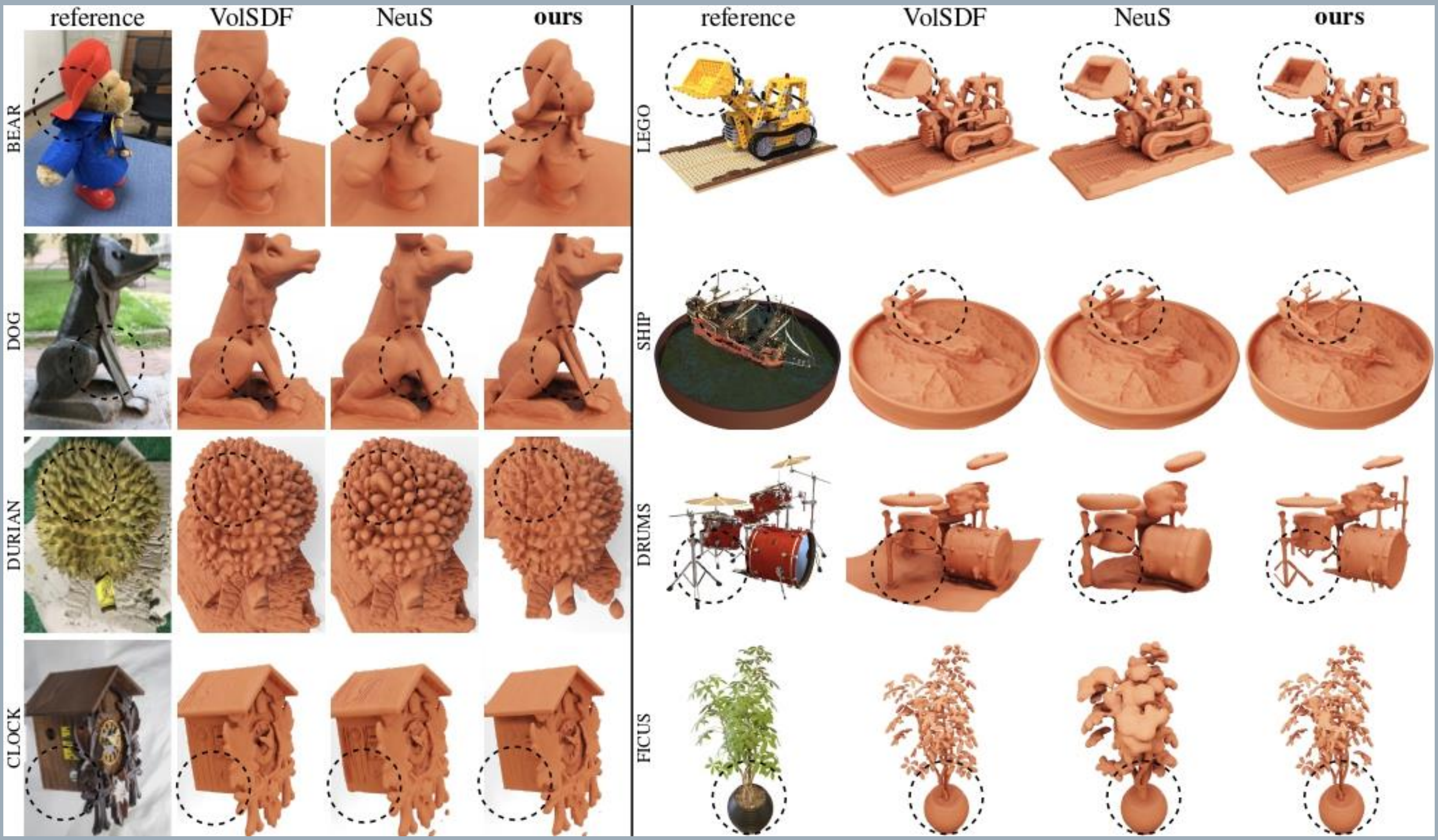
CVPR DAILY

Seattle 2024

Computer Vision and Pattern Recognition

Wednesday



Awards, Highlights, Challenges, Workshops, Previews, Women in Computer Vision, expressly reviewed for CVPR 2024!

In cooperation with

Computer Vision News

The Magazine of The Algorithm community

A publication by

RSIP VISION

Good morning, CVPR!

Welcome to the first edition of **CVPR Daily for 2024**! **CVPR** and **RSIP Vision** (the publisher of **Computer Vision News**) have partnered again to bring our daily magazine to the community throughout this incredible event. Follow us today, tomorrow, and Friday, and let us help you navigate all the innovative content this week.

Can you feel the buzz in the air? We're thrilled to announce a record number of **11,600 participants this year, with around 80% joining us in Seattle**. We're back, baby! We've seen the quality of the accepted papers and can promise you're in for a treat. The rise of **multimodal AI** means we're starting to see a convergence of various modalities in language and vision. With such intense emphasis on AI right now, huge corporate interest, and investment from academic labs in vision-based topics, there's truly no better time to be here!

This year, you must check out our expanded **Art Program**, which includes an AI art gallery created by the renowned curator, producer, and researcher Luba Elliott. This big space features interactive installations and opportunities to meet and hear from the artists. Luba will also conduct gallery walkthroughs, offering an entertaining way to experience this new application of computer vision.

Our **Expo, featuring over 100 companies**, is always a popular part of the conference and is another must-visit this year. Companies from all areas of computer vision are here to showcase cutting-edge products and search for new talent. As an industry-friendly academic field, networking at the Expo can be just as valuable as networking at the poster sessions.

Beyond the fascinating talks and posters, we've been working on expanding the program considerably. We've planned socials to bring different affinity groups together. Our big conference reception, featuring music by our attendees, is an excellent opportunity to relax and meet people. We've just seen a vast slate of workshops and tutorials to bring together those working on the same research topics. We'll keep thinking of new ways to add this kind of programming, so watch this space!

If you're new to CVPR, don't feel out of place if you don't know anyone yet. Everybody here was new at some point, and many lasting connections are made during this event**. Introduce yourself, join the social events, and engage in conversations**. You're here with a community – you're not alone!

We encourage everyone to attend the **PAMI-TC meeting** on Thursday at 4pm. This open session is dedicated to the business of our community, offering insights into key issues and providing a platform for getting involved. All are welcome, and everyone can vote on the motions.

As **General Chairs**, we're proud to endorse this wonderful magazine and hope you enjoy reading the first issue. Please share it with your friends and colleagues, and don't forget to come back tomorrow and Friday morning for more. **We wish you all a successful CVPR!**

General Chairs:

Octavia Camps (Northeastern University)
Rita Cucchiara (University of Modena and Reggio Emilia)
Walter Scheirer (University of Notre Dame)
Ramin Zabih (Cornell University)
Sudeep Sarkar (University of South Florida)

... with Ralph Anzarouth (CVPR Daily and Computer Vision News)

CVPR Daily

Publisher: **RSIP Vision**
Copyright: **RSIP Vision**

Editor: **Ralph Anzarouth**

Our editorial choices are fully independent from IEEE, CVPR and the conference organizers.

Objects as Volumes: A Stochastic Geometry View of Opaque Solids

Bailey Miller is a PhD student at Carnegie Mellon.

His novel paper, which breaks new ground in volume rendering, has been selected from thousands of accepted papers as a conference highlight and is in the running for a coveted Best Paper Award.

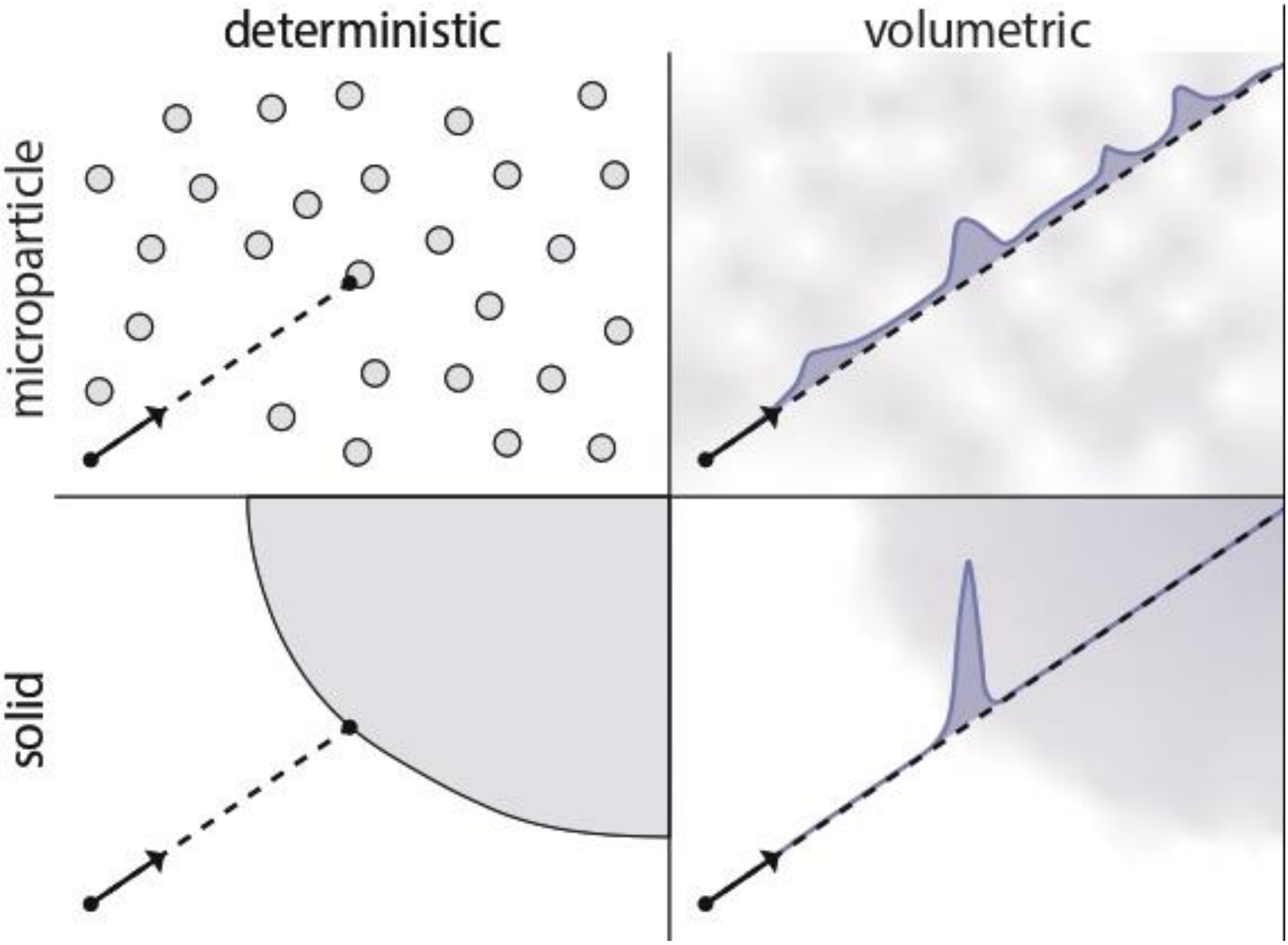
Bailey speaks to us before his oral presentation this morning.

For several decades, **volume rendering techniques** have been a popular class of methods for simulating light transport in translucent media such as clouds, smoke, and tissue, with various applications in graphics and physics. In the past five years, there has been a shift towards using these methods to model more familiar, everyday objects such as solid, opaque items.

"Our paper is about figuring out why these volume rendering methods, originally developed for clouds, can work on things like a Lego truck," Bailey begins. *"**We've developed a stochastic geometric theory that explains the connection between these two different models.**"*

The initial challenge Bailey faced was understanding the foundational principles of volume rendering, which have been obscured or had a black box put around them by the numerous successful yet complex methods developed in recent years. *"Revisiting its roots, you see that in classic volume rendering, scenes are modeled as a collection of microparticles,"* he reveals. *"Once we could understand it in this very principled manner, we could start to develop **ideas and approaches for considering volume rendering on stochastic opaque solid objects.**"*

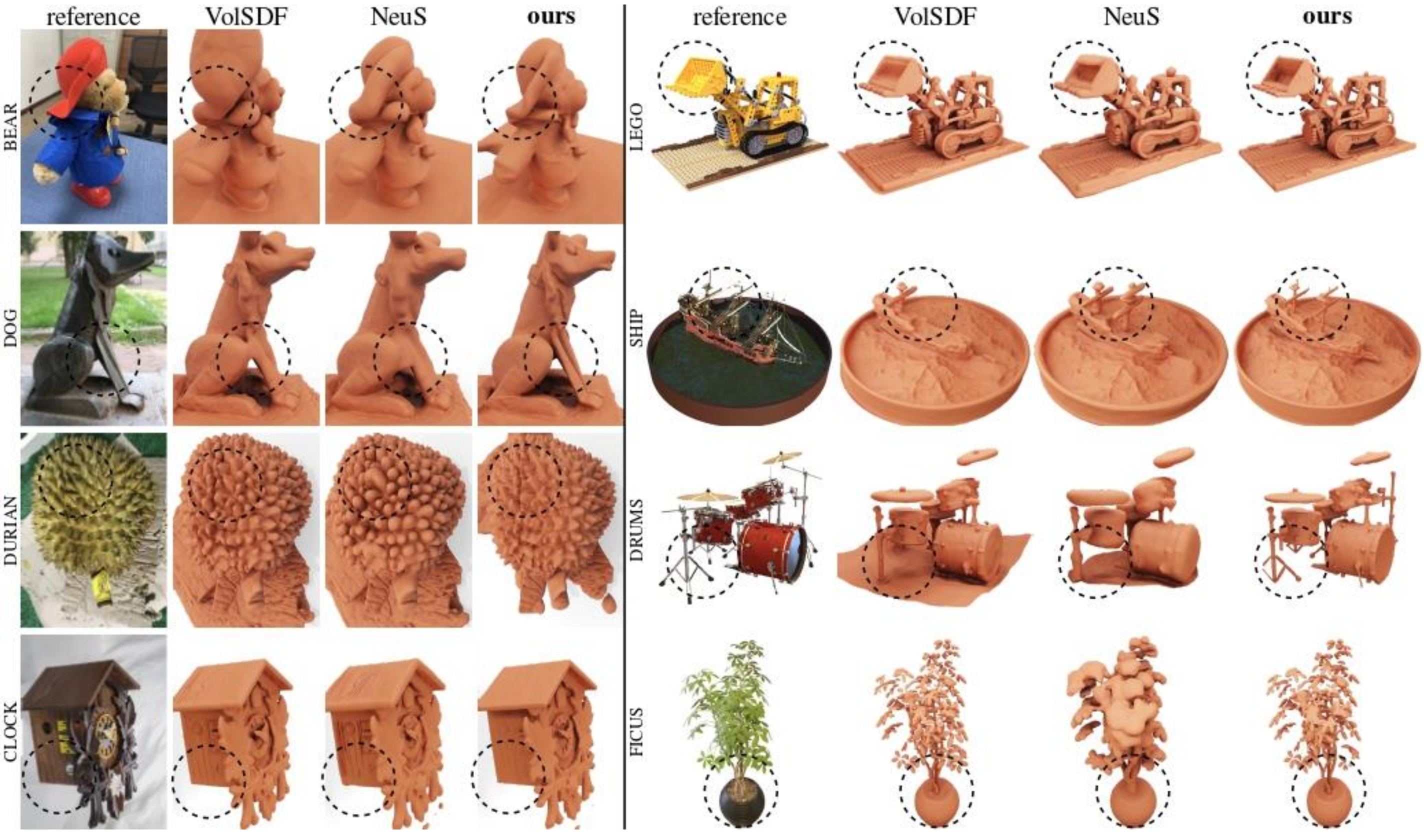
Did he solve the problem? *"Part of it,"* he tells us. *"I think we've opened some new doors. We show how you can develop these rendering algorithms for a very*



limited set of new stochastic geometry. There's a lot of work to be done in extending these methods to even more extensive types of geometry and scenes."

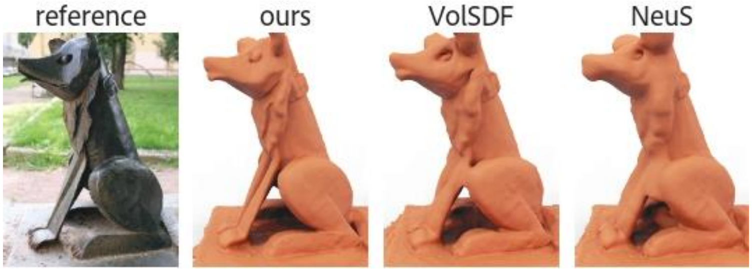
One of the primary applications of this work is in **surface reconstruction**. Essentially, this involves taking a collection of images and trying to understand the geometry of the world that gave rise to them. The connection becomes clearer through light transport. *"Light bounces around the world, and depending on how it interacts with the geometry, it gives different images,"* Bailey explains. *"By introducing a new way of modeling the geometry in scenes and considering how light interacts with that geometry, we can improve the surface reconstruction algorithms that have been using volume rendering over the past several years* ***to get all sorts of performance improvements****."*

The implications of this research extend to various fields, particularly settings such as **robotics** or **autonomous driving**, where there is a benefit in having a notion of uncertainty in solid objects and the



trustworthiness of algorithmic results. **In these scenarios, agents might leverage images or video to build a probabilistic model of the surrounding world, which can significantly enhance reliability, safety, and efficiency**.

Being chosen as an **award candidate for the first time** is a high enough achievement, but it is even more special for Bailey, as **this is his first CVPR**. He puts the positive reception down to how timely the work is given recent advancements in surface reconstruction and novel view synthesis algorithms based on volume rendering, which have worked well but for unclear reasons. *"By providing this perspective and saying, 'Here's an explanation for why all these volume rendering methods work so well,' we can understand why we can model the world in this particular way and why we've had success with it,"* he points out. *"We can understand the current state of the art but also develop a perspective on these methods that allows us to continue improving. I love thinking about how we model the world and how the underlying assumptions we make in our models impact the algorithms and methods we ultimately develop."*



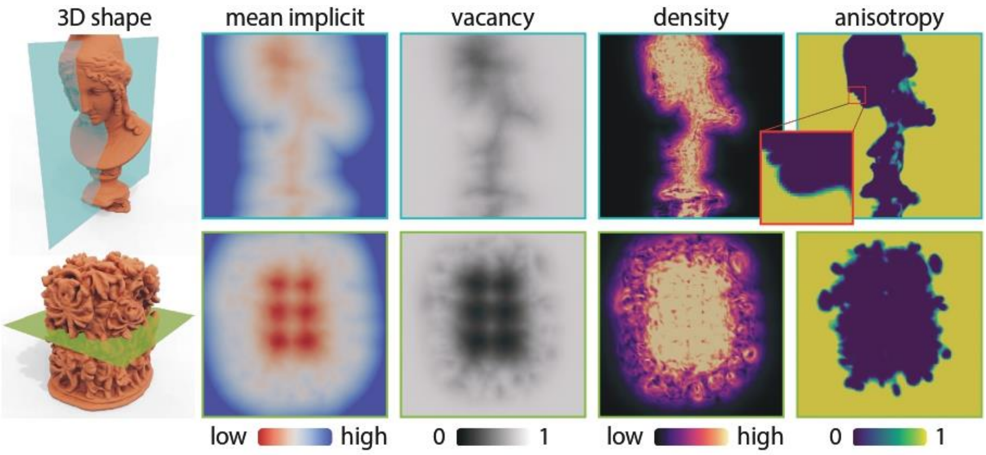
Reflecting on the growing emphasis on the role of **explainability in research**, Bailey says it is a trade-off: *"You need to push forward and figure out what works in practice, and then you need to step back and ask, 'Why do these methods work so well?' It's a constant process moving back and forth between the two."*

In addition to his work on stochastic geometry, Bailey is exploring **Monte Carlo PDE solving**, which involves adapting Monte Carlo methods that work really well for light transport and simulating light to other types of physics, such as **heat transfer, acoustics, and wave equations**. *"I don't think this has been as present in the vision community yet, but it's been starting to gain some attention in graphics,"* he tells us. *"I think, eventually, these algorithms will be of interest in the computer vision community because we're seeing the development of all sorts of new imaging modalities or renewed interest in modalities like* ***thermal imaging****. Good ways to simulate those should help vision researchers and practitioners* ***develop algorithms that work with physics beyond just light****."*

Looking ahead, **Bailey is excited about the potential for developing this work further, including extending the stochastic geometric approach to a broader range of stochastic models and probabilistic assumptions about the world or scenes**. Also, the core idea of stochastic geometry has applications beyond light transport algorithms, which opens a range of possibilities for future research.

Could we be sensing the first hints of next year's award paper? *"I'd love that, but I'm happy with the one this year for now!"* he laughs. *"I feel very fortunate to have had our paper selected. I hope everyone who reads it enjoys it and takes something away from this stochastic geometry perspective."*

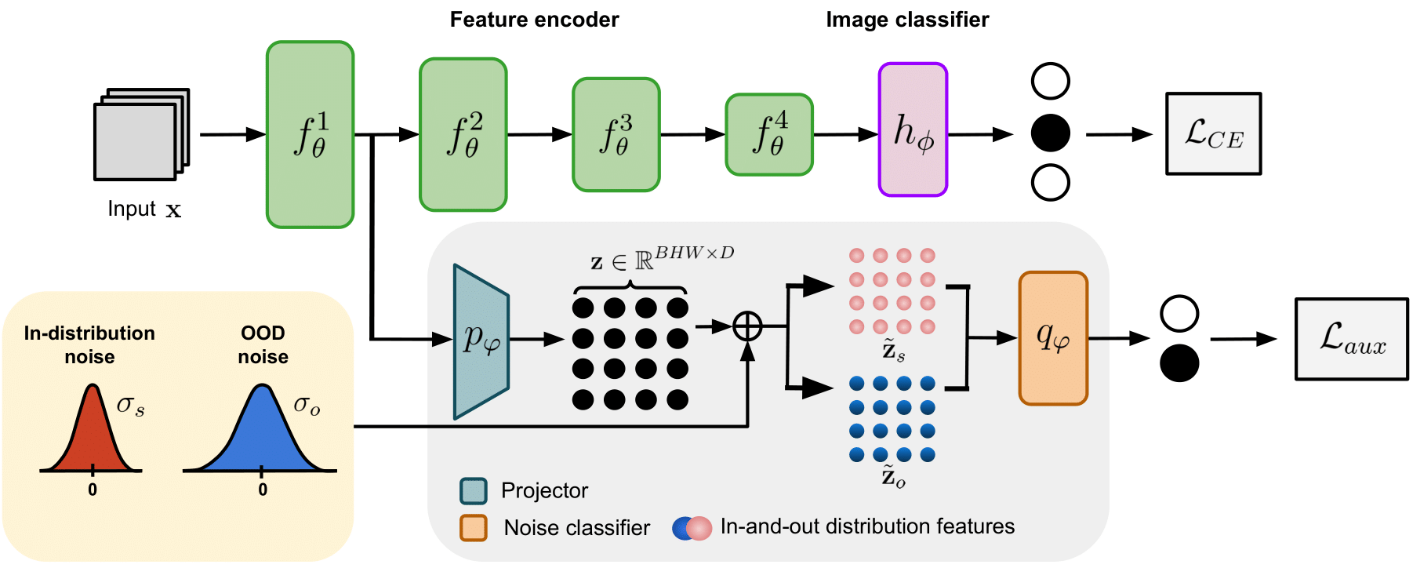
To learn more about Bailey's work, visit Oral Session 1B: Vision and Graphics (Summit Flex Hall AB) from 9:00 to 10:30 [Oral 4] and Poster Session 1 & Exhibit Hall (Arch 4A-E) from 10:30 to 12:00 [Poster 412].



NC-TTT: A Noise Contrastive Approach for Test-Time Training

David Osowiechi (left) and Gustavo Vargas (right) are PhD students at ETS Montréal. Their supervisors and co-authors are Christian Desrosiers and Ismail Ben Ayed. Their work on test-time adaptation has been picked as a highlight paper, and David and Gustavo speak to us before their poster session this afternoon.

While deep learning models have demonstrated outstanding performance in vision tasks, they often struggle when confronted with domain shifts at test time – where the test data differs significantly from the training data. **Test-Time Training (TTT)**, a specific

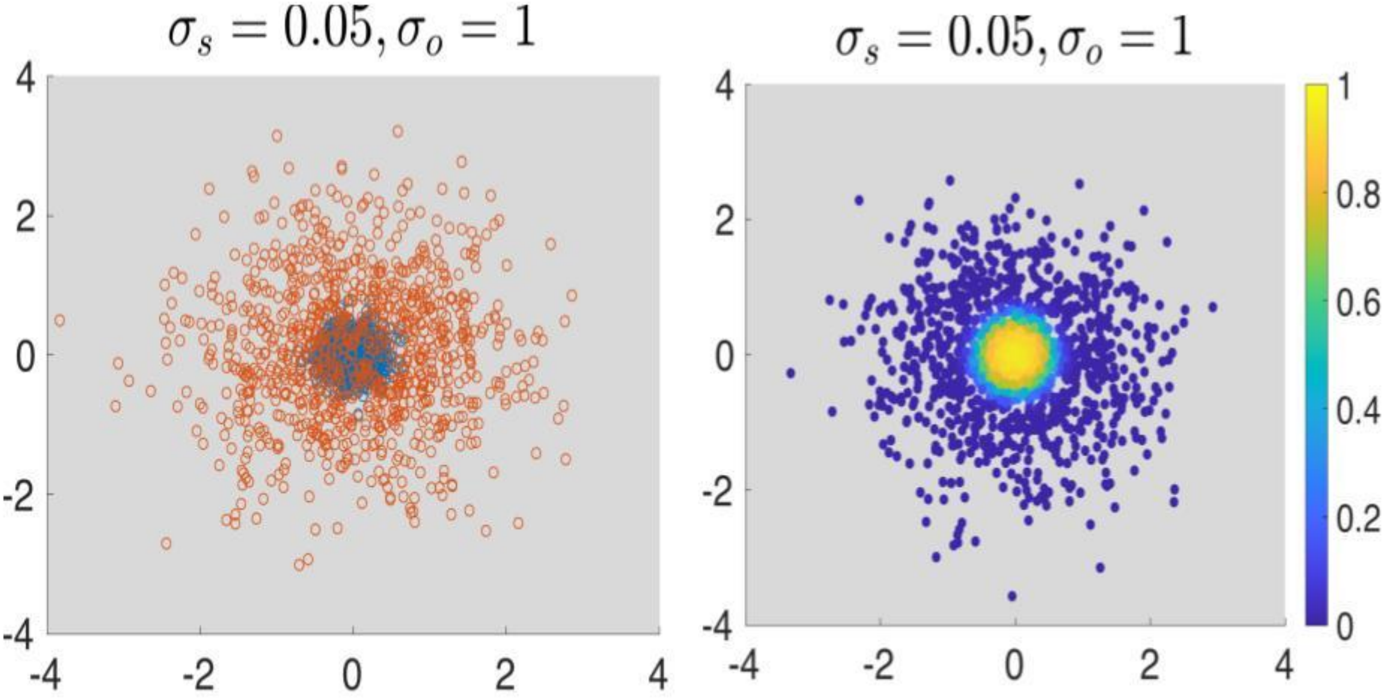


approach within the broader category of **Test-Time Adaptation (TTA)**, has emerged as a promising way to address this challenge and enhance model robustness.

"In TTA, we take a model that has been pre-trained on a source dataset and adapt it on a target dataset at test time, but ***without labels and images from the source dataset****,"* David explains. *"To do this, there's an architecture named TTT, where we introduce an auxiliary task at test time,* ***which is often self-supervised or unsupervised, without labeled data****. At training time, we train the model, and at test time, we update the feature extractor thanks to this auxiliary task,* ***which improves the model's performance****."*

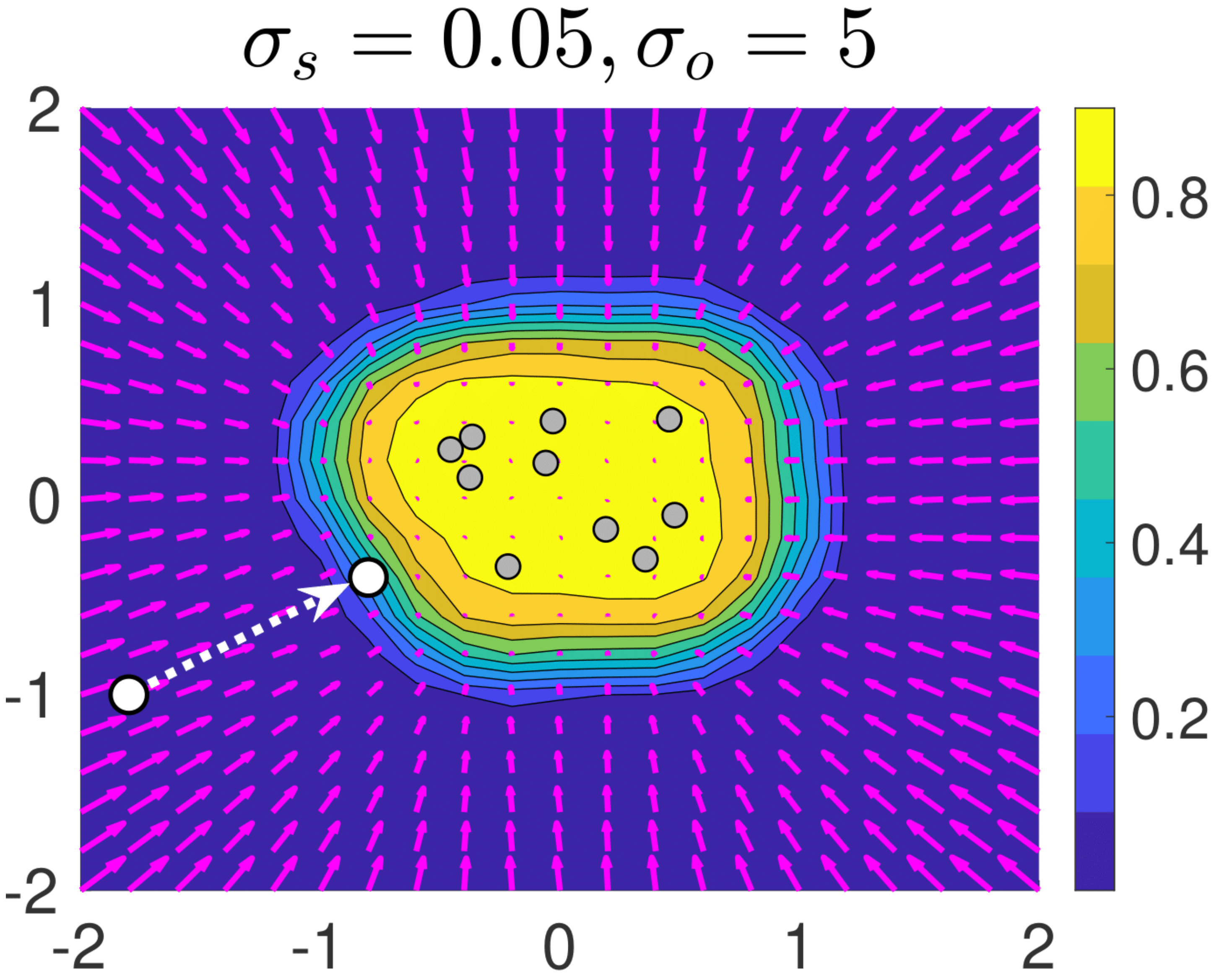
David and Gustavo introduce the concept of **Noise-Contrastive Test-Time Training (NC-TTT)**, a novel unsupervised TTT technique based on the popular theory of **Noise-Contrastive Estimation**. With NC-TTT, the model learns to classify noisy views of projected feature maps and then adapts accordingly on new domains. It employs the same Y-shaped architecture as previous works, with one branch dedicated to the main classification task and the other to the auxiliary TTT task.

NC-TTT is particularly suited for classification tasks. However, the principles of TTA can be applied to other vision tasks, particularly those prone to **domain shifts caused by noise or other corruptions**. Even



changing from black and white to color can represent a domain shift for a model trained in black and white. “*For an easy example, say you train your model with images of cats and dogs, and you want to classify it,*” David illustrates. “*You’ve only taken pictures at training during the spring, so then at test time, you’ll show some images that you took in the winter, and your model’s performance will decrease because it never saw images in winter. The goal of our architecture is to adapt to the winter images because it already has this knowledge at spring.*”

Gustavo tells us that TTA should be an essential component of any model deployed in the real world because **models often encounter images that differ from their training data**. “*There will always be images that might be slightly different, or the camera that’s used to get the new images will change,*” he points out. “*At some point, the model’s performance will decrease, and the only way to ensure the model keeps going well is to have an adaptation mechanism that can work in real time or, in this case, at test time.*”



The team's challenge was to **surpass state-of-the-art performance with their new method**. The paper demonstrates that their approach can recover classification performance significantly compared to existing TTA baselines.

Noting that this is still a relatively new field, David and Gustavo are eager to share their work with the broader research community at **CVPR** today. Their poster session is an opportunity to introduce NC-TTT and foster discussions about the future of TTA and TTT in general. *"I'd invite any deep learning researcher to think about how a model they train can fail at test time,"* Gustavo urges. *"This is just one alternative we can provide to prepare the model before that happens.* ***The most exciting part is coming up with new auxiliary tasks that work better and better!****"*

Looking ahead, David sees their work as **the beginning of a broader exploration of TTA**. *"It's not the end of something,"* he surmises. *"We've seen a lot of TTT and TTA articles. Many fields are progressing in TTA,* ***like working on Vision-Language Models (VLMs) or segmentation****. I think it's something that will be everywhere in the future – this is just the beginning."*

To learn more about David and Gustavo's work, visit Poster Session 2 & Exhibit Hall (Arch 4A-E) from 17:15 to 18:45 [Poster 123].

Who is our Woman in Computer Vision today? Find out on page 24!

NRDF: Neural Riemannian Distance Fields for Learning Articulated Pose Priors

Yannan He is a PhD student with the Real Virtual Humans Group within the Department of Computer Science at the University of Tübingen, supervised by Gerard Pons-Moll.

His work, selected by area chairs to be a highlight paper, follows on from second author Garvita Tiwari's award-winning work, Pose-NDF.

Yannan speaks to us before his poster session this morning.

The seeds for this project were sown when Yannan and the team behind **Pose-NDF**, a continuous model for plausible human poses based on **neural distance fields**, began to investigate the failure cases of Pose-NDF. They discovered an underlying issue in its **training data distribution**, which should show decreasing samples as you move away from the pose manifold. In this new paper, NRDF, Yannan aims to set that right.

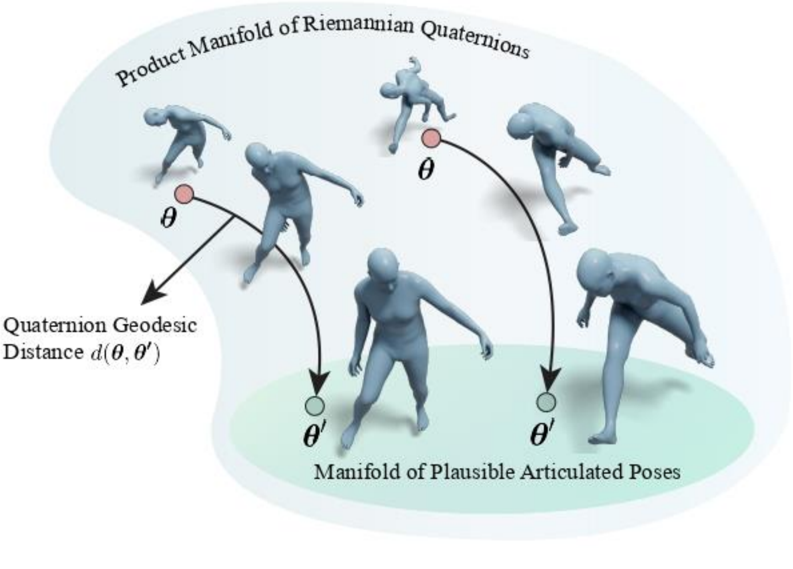
"We're still modeling it as a distance field," he explains. *"During inference, it could help if you draw more samples nearby the manifold because, during the projection, you're moving closer and closer to the manifold. If there are more negative samples near the manifold, the network prediction will be more accurate, and it'll help you* ***achieve more stable and accurate projection results****."*

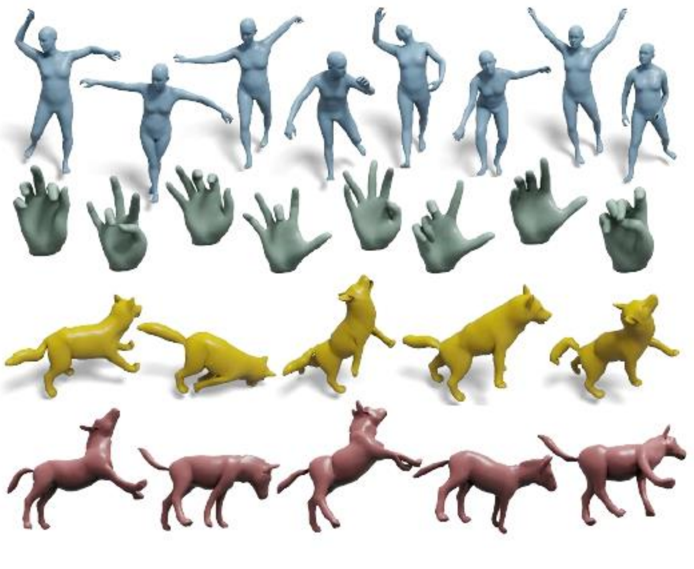
The goal is to model distance field-based pose priors in the space of plausible articulations. The learned pose priors are versatile and can be applied to various downstream tasks such as pose denoising, 3D pose estimation from single images, and solving inverse kinematics from sparse observations. *"Sometimes existing methods return 3D poses where the image overlay is good, but if you view it from another point of view in 3D, it's implausible,"* Yannan points out. *"The 3D pose itself may have self-occlusions, interpenetrations, and also some implausible pose patterns, like a knee bending outwards."*

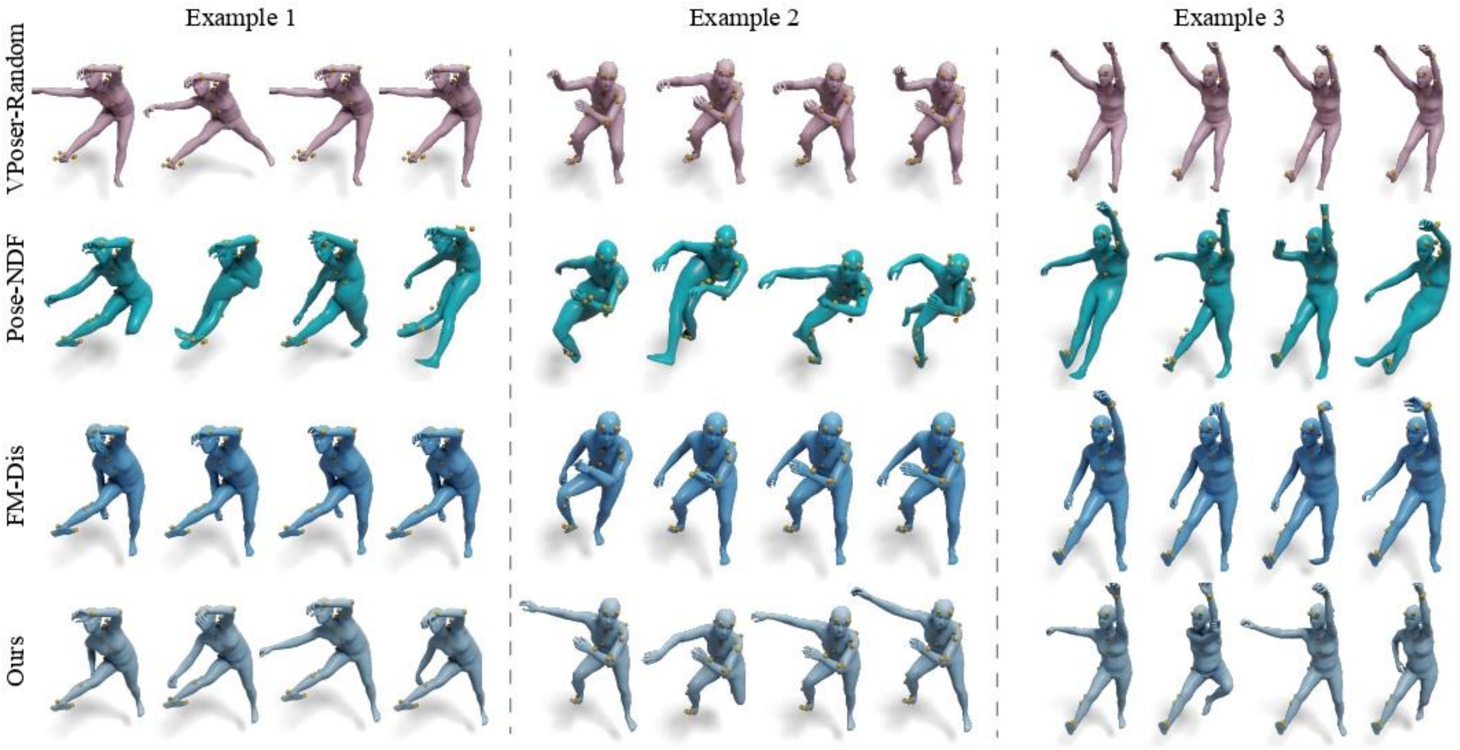
Besides humans, NRDF can be extended to any articulated shapes, such as hand and animal poses. It can return plausible and valid results with only wrists and ankles as surface markers.

Yannan outlines two critical ideas presented in NRDF. Firstly, **a novel sampling approach to address inaccurate distance prediction**, where there are no training samples near the manifold. *"During the training data preparation, we aim to draw more samples near the surface, with a gradual decrease as we move to the faraway regions,"* he explains. *"In the distribution of Pose-NDF, there is a huge gap between the zero-level-set and the mean, which lies in the center with a big distance value. After we propose the sampling algorithm, we could obtain a distribution shape like a half-Gaussian distribution. Actually, it could be any distribution that the user specifies. It could be an exponential distribution or a uniform distribution."*

Most distance fields work with 3D point clouds, so they are learning 3D shapes in Euclidean space, but this work features **high-dimensional articulated poses represented by K quaternions in the product space of SO(3)**. *"Sampling points in Euclidean space is totally different from directly sampling rotations,"* Yannan tells us. *"The special thing about our work is we propose an easy way to*


Product Manifold of Riemannian Quaternions
θ
θ
Quaternion Geodesic Distance $d(\theta, \theta')$
θ'
θ'
Manifold of Plausible Articulated Poses

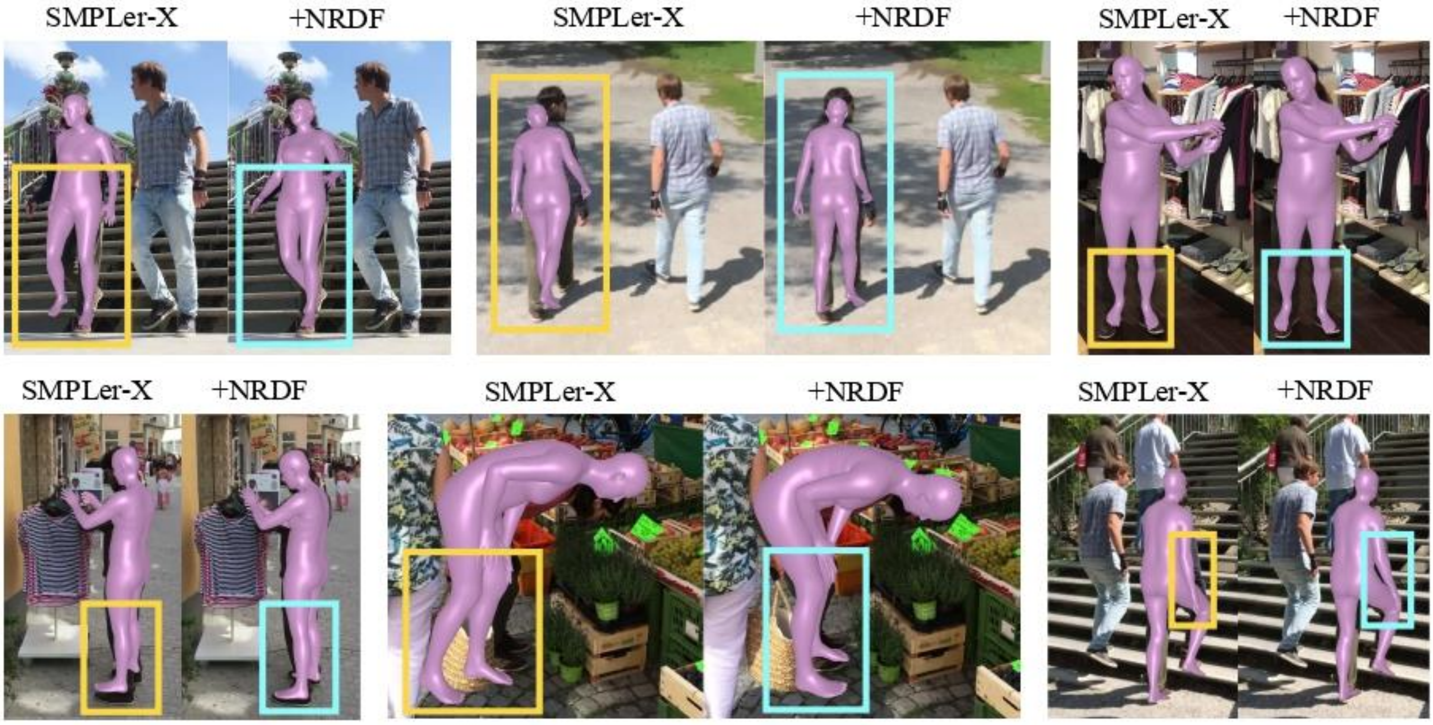




directly sample articulated rotations in the articulated SO(3) space, which we call ***a product manifold of Riemannian quaternions****."*

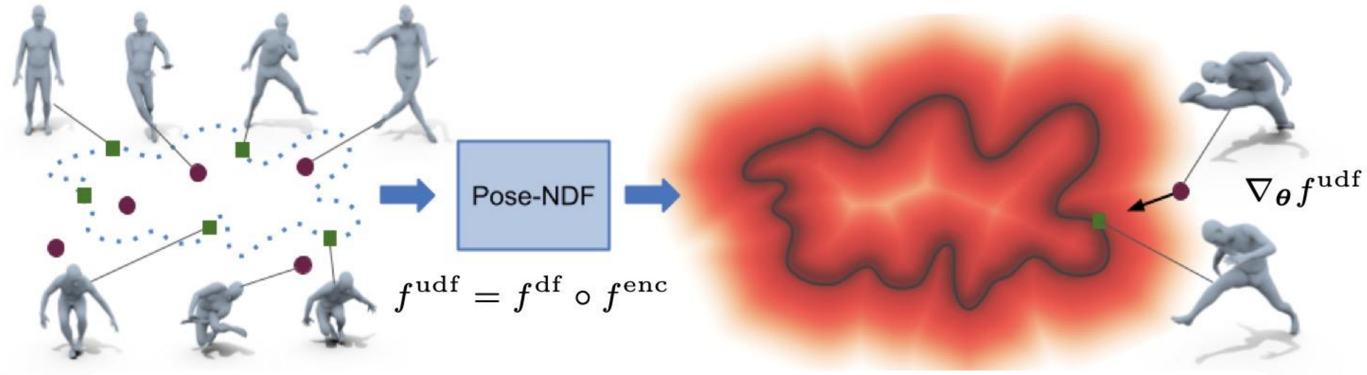
Additionally, this work introduces **RDFGrad, an innovative technique that streamlines the gradient descent process during inference**. For Pose-NDF, after a normal gradient descent, you have to reproject the resulting pose onto the quaternion space because the quaternion should also be uniform, which slows down the projection process. In contrast, NRDF extends the original gradient descent procedure onto the Riemannian manifold during inference time projection. *"Given a noisy pose, we obtain the gradient direction returned by the network propagation, which is the Euclidean gradient, and we iteratively project it onto the tangent space of a given pose and directly work along the geodesic"* he explains. *"This is crucial and makes the process faster."*

Yannan tells us the breakthrough came with the realization that **optimizing the training data distribution is crucial for distance fields. Finding this, out of all possible explanations for Pose-NDF limitations, was one of the biggest challenges and took several months of intense investigation. For NRDF, no network architecture** was modified compared with Pose-NDF.



Collaborating with Gerard Pons-Moll and a talented team of researchers was a rewarding experience for Yannan. *"It's a lot of fun working with Gerard because he's a really nice person. Even away from research, I can learn a lot from him – he's a big fan of music, and I also enjoy making music!"* he laughs. *"Also, our other collaborators,* ***Jan Eric Lenssen****,* *Tolga Birdal, and Garvita Tiwari. We're diving deeper into the mathematics behind the problem and exploring the essential reasons behind the phenomenon. It's really cool, and I learned a lot from them."*

To learn more about Yannan's work, visit Poster Session 1 & Exhibit Hall (Arch 4A-E) from 10:30 to 12:00 [Poster 145].


Pose-NDF
$f^{\text{udf}} = f^{\text{df}} \circ f^{\text{enc}}$
$\nabla_\theta f^{\text{udf}}$

Garvita Tiwari's work Pose-NDF that won a Best Paper Honorable Mention award at ECCV 2022.
Read our full review.

Russian Invasion of Ukraine

CVPR condemns in the strongest possible terms the actions of the Russian Federation government in invading the sovereign state of Ukraine and engaging in war against the Ukrainian people. We express our solidarity and support for the people of Ukraine and for all those who have been adversely affected by this war.

CVPR
JUNE 17-21, 2024
SEATTLE, WA
Thank You to our Corporate Sponsors
Platinum Sponsors
amazon | science
AMD
captions
Google
Hyperbolic
Lambda
Meituan
Meta
Qualcomm
SONY
Gold Sponsors
Structured Semantic 3D Reconstruction Challenge:
1st Place Solution and Takeaways
Denys Rozumnyi
Computer Vision and Geometry Group, Department of Computer Science, ETH Zurich
Visual Recognition Group, Faculty of Electrical Engineering, CTU in Prague

Denys Rozumnyi has won the Structured Semantic 3D Reconstruction Challenge, which was held as part of the Urban Scene Modeling workshop here at CVPR 2024.

The objective of this competition is to facilitate the development of methods for transforming posed images into a structured geometric representation, from which semantically meaningful measurements can be extracted. In short: More Structured Structure from Motion.

CVPR 2022
DAILY
Computer Vision and Pattern Recognition
Thursday
Workshop: Computational Cameras and Displays
Presenting Work by: Ruslan Partsey, Ozan Özdenizci, Dima Damen & team
Exclusive Interview with: Mathieu Salzmann
Women in Computer Vision: Anna Rohrbach
Today's Picks by: Anh N. Thai
In cooperation with Computer Vision News
The Magazine of The Algorithm community
A publication by RSIP VISION
Tuesday
zebra
Workshop: Medical Computer Vision
Presenting Work by: Bowen Cheng, Ping Hu, Fabio Cermelli and Dario Fontanel
Editorial with: Program Chairs
Women in Computer Vision: Shuran Song
Today's Picks by: Maria Dobko
Wednesday
False Positive
Friday

Seamless Human Motion Composition with Blended Positional Encodings

German Barquero (left) is a third-year PhD student working on human motion understanding and generation under the supervision of Sergio Escalera and Cristina Palmero (right).

Cristina is a freelancer in computer vision and machine learning applied to human behavior, understanding, and synthesis. She is collaborating with the University of Barcelona on a project exploring an important problem in human motion generation.

German and Cristina speak to us before their poster session this morning.

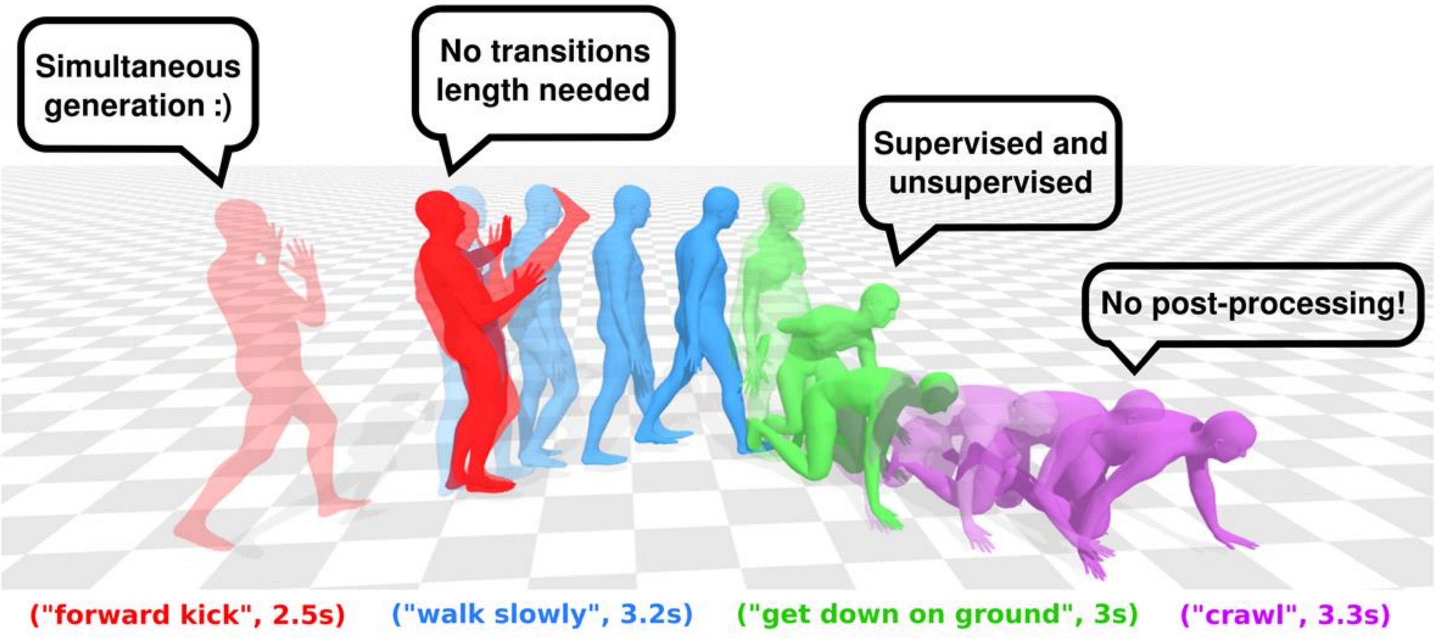
Traditional methods in **human motion generation** have focused mainly on isolated, short-duration motions, often guided by text, music, or scenes. While innovative, **these techniques fall short when producing long, continuous sequences**, seamlessly transitioning from one motion to the next.

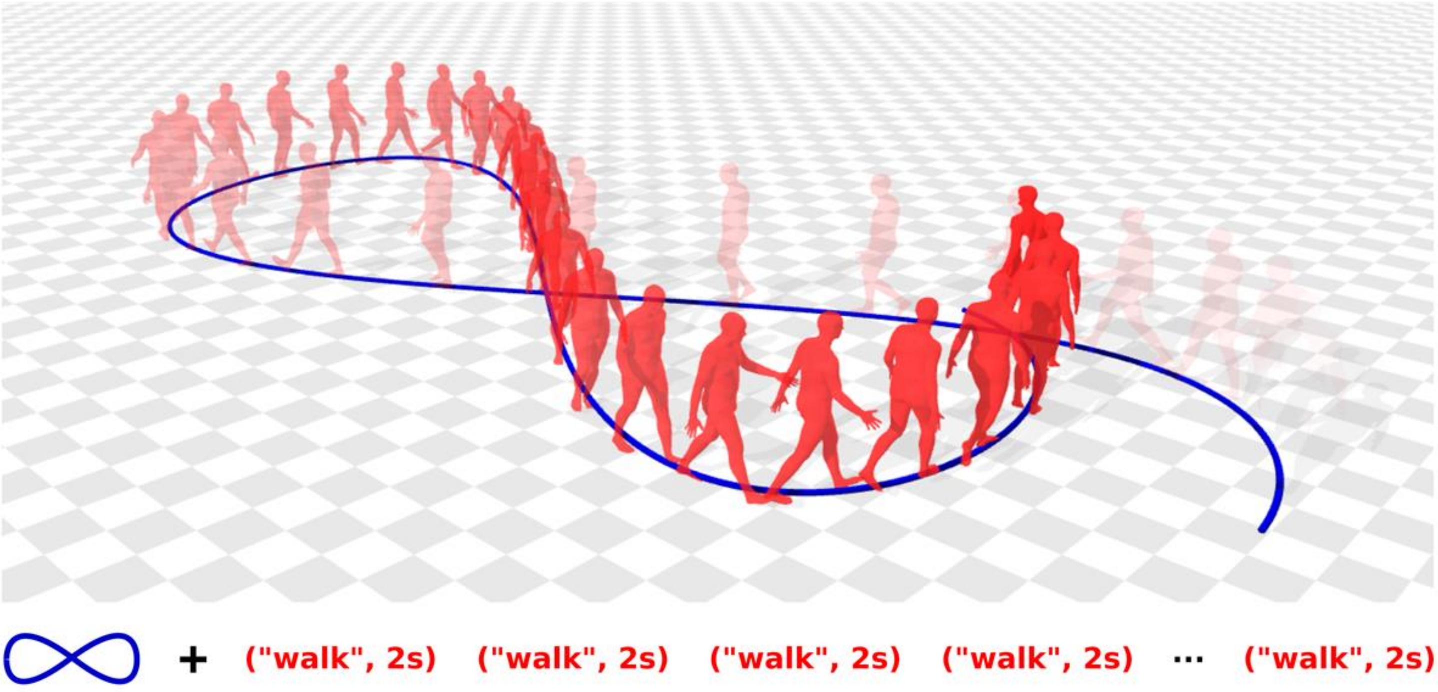
This paper takes a fully learning-based approach to generate motions according to multiple actions in a sequence. It proposes **FlowMDM**, the first diffusion-based model that **generates seamless human motion compositions guided by textual descriptions without postprocessing or extra denoising steps**.

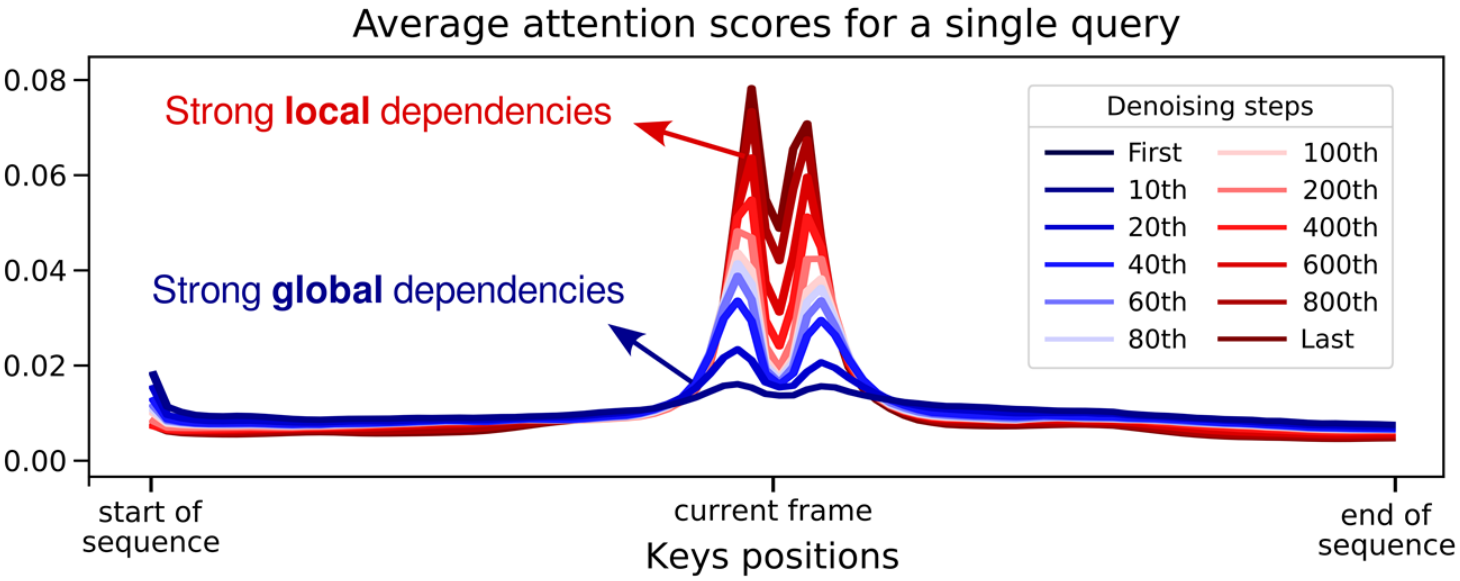
"We're really proud of this paper," German asserts enthusiastically. *"It addresses a very important problem in human motion generation that we*

think was being ignored. Prior works used a model to generate sequences of different motions ***according to different actions from an engineering perspective****, generating each action individually and* ***then applying postprocessing*** *in the transitions to make a motion transition emerge.* ***We do it with learning but don't apply postprocessing in the transition.****"*

"The generated human behavior compositions are more realistic than ever," Sergio adds. *"Compared to previous work we can see a clear improvement in accuracy, realism and smoothness. Its application benefits of only requiring one condition for training and generalizing to several inference conditions, without requiring any postprocessing!"*


Simultaneous generation :)
No transitions length needed
Supervised and unsupervised
No post-processing!
("forward kick", 2.5s)
("walk slowly", 3.2s)
("get down on ground", 3s)
("crawl", 3.3s)

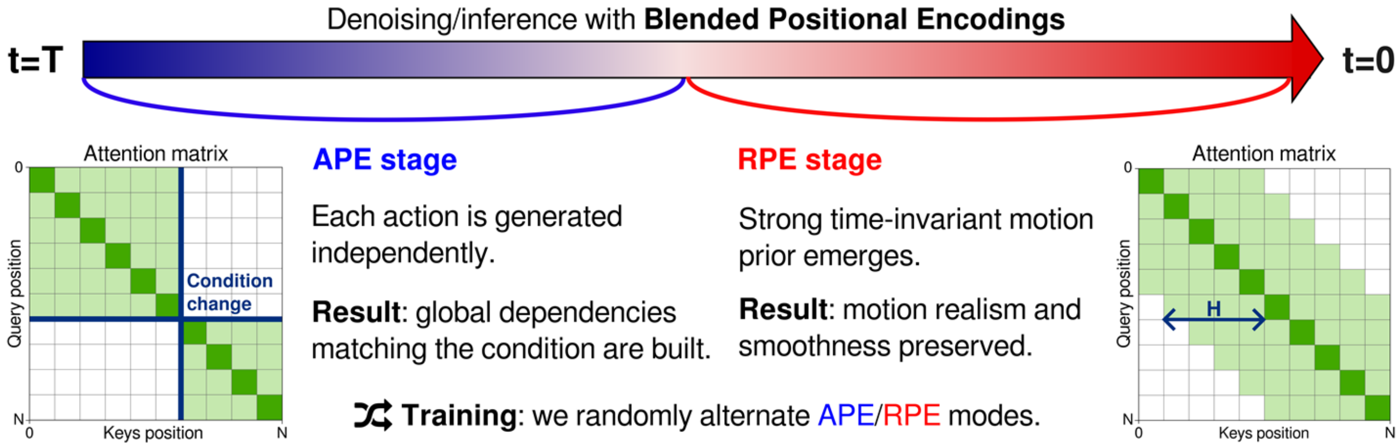

∞ + ("walk", 2s) ("walk", 2s) ("walk", 2s) ("walk", 2s) ··· ("walk", 2s)



Gaming, filming, and animation are typical applications for FlowMDM, where text is a very natural, powerful, and user-friendly way to guide motion generation and control motion semantics. *"Let's say you're preparing a storyboard and imagining how things will look,"* Cristina says. *"You can use text-to-motion. First, a person will start walking, and then they will kick a ball with their right foot. You can see this motion.* ***If you can visualize things, you can better appreciate if you like them or not.****"*

FlowMDM has potential applications in many other areas, including **human-computer interaction, physical training, and robotics**. *"We could be guiding a robot in surgery, for example,"* German points out. *"We know that it needs to extract an organ, and we need to generate the transitions between extracting the organ and putting it in someone else.* ***Everything is guided by generating transitions****. It's a really central problem."*

The team's greatest challenge was



APE stage

Each action is generated independently.

Result: global dependencies matching the condition are built.

RPE stage

Strong time-invariant motion prior emerges.

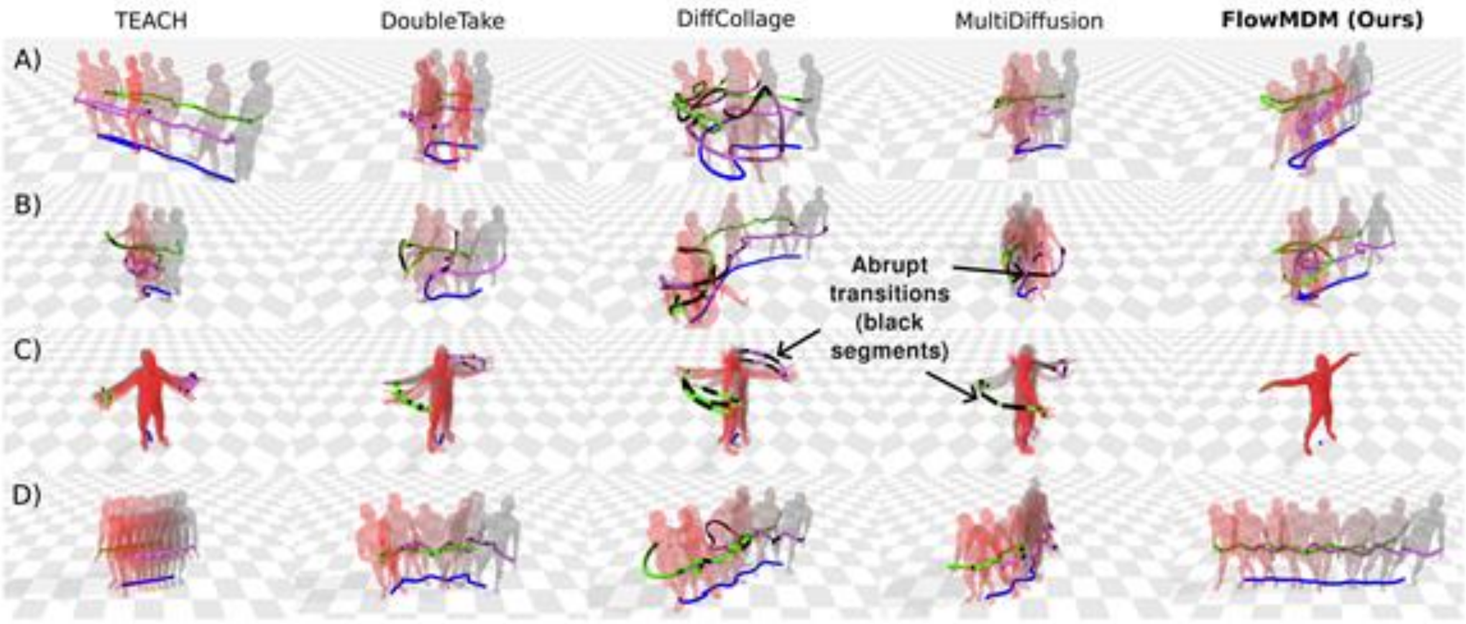
Result: motion realism and smoothness preserved.

Training: we randomly alternate APE/RPE modes.

sticking to its very clear goals, including **avoiding the need to determine transition lengths manually, which can be ambiguous and subjective**, and not applying any postprocessing. *"We observed in prior works that if you closely analyzed the motions they were generating, in the boundaries of this postprocessing, they showed some **artifacts and abrupt transitions**,"* German continues. *"These were a result of this artificial postprocessing they were applying, and we didn't want that."*

To transfer the semantics of the text into the motion, **the model needs to know the absolute position of each frame within the sequence**. *"Let's say we want to generate five minutes of motions of different actions sequentially,"* German suggests. *"We need to know the beginning and ending of each action. Also, we want it to be invariant to the absolute position because we don't have training data that takes as long as five minutes. We want a method that can generate motion, no matter at which position, which is against having access to the absolute position. We have two things fighting against each other and need to harmonize them somehow."*

A breakthrough came in the form of **Blended Positional Encodings (BPE)**, a new concept for diffusion models that uses **both absolute and relative positional encodings in the denoising process**. In essence, diffusion models transform noise into the target motion. **Initially, absolute information is used to recover the global motion coherence, and then relative positions are used to build smooth and realistic transitions between actions**.



This innovative work analyses a problem common in many fields, not just motion but image and video generation and anything related to generation with diffusion models. It exploits the denoising dynamics of diffusion models to guarantee particular conditions in downstream applications. *"In deep learning and computer vision, the aim is to make models learn effectively from the data provided,"* German explains. *"This research is intriguing because* ***it introduces specific inductive biases into the network. These biases help the network improve at certain tasks****. In this case, generating smooth transitions between actions without being explicitly trained on those transitions. This approach is important for researchers as it enhances the network's ability to perform desired tasks efficiently."*

Indeed, the team is eager to test its methods in other scenarios beyond motion generation, where perhaps the constraints are not the same as in the motion domain but need further exploration. *"This work solves a specific problem very effectively and simply,"* Cristina asserts. *"That's very important. Trying not to complicate things and using this end-to-end method, I believe it can be a very good contribution to the field."*

To learn more about the team's work, visit Poster Session 1 & Exhibit Hall (Arch 4A-E) from 10:30 to 12:00 [Poster 28].

Who is our Woman in Computer Vision today? Find out on page 24!

June 2024

Computer Vision News

& Medical Imaging News

The Magazine of the Algorithm Community

ICLR
H INTERNATIONA ONFERENCE ON LE ING REPRESENTA
Vi , Austria | May 7-11, 2024
ICLR
ICLR
LR

Did you read Computer Vision News of June?

Read it here

Special ICLR Edition with full reviews of award-winning papers! Pages 2 to 33

A publication by RSIP VISION

"Now suddenly, everything is converging into one big model, which can do these various things, which is exciting but also maybe a bit scary!"

Ivana Balažević is a Research Scientist at Google DeepMind.

She spoke to us at CVPR 2024 on Monday, right after her talk and panel at the Prompting in Vision workshop.

Read 160 FASCINATING interviews with Women in Computer Vision

Where does Balažević come from?

Croatia.

I am not very far away. I am Italian.

Ah, okay, so we're neighbors! [*she laughs*]

What is your work about?

Well, various different things. I have mainly, in the past couple of years, worked on multimodal image and video understanding. As of recently, I moved into Gemini, working more on language. But, yeah, Gemini, super secret, can't talk about it – you know how it is!

Is convergence of multimodalities really happening? Text with vision, video, audio, all these things together?

I think it is, especially in the past couple of years. I finished my PhD in 2021, and there were just these small models doing various different tasks. Everyone was working on their little model in their PhD or in whichever company, and now suddenly, everything is converging into one big model, which can do these various things, which is exciting but also maybe a bit scary. I don't know. Mainly exciting, I would say.

Why exciting, and why scary?

That's a very good question! In my mind, exciting because it unlocks a whole world of possibilities for what we can possibly do with these models in some possibly distant future. I don't know because I didn't think we'd be where we are now, but we would maybe be able to learn from these models or learn something new that we don't know. These models might be able to make some sort of inferences, like combining various modalities to teach us things. How amazing would it be if we had a model that would be able to read scientific papers and come up with a new paper that is actually correct and that teaches us something new? Or a model that takes all our knowledge about medicine and biology and chemistry and finds a cure for cancer or something like that? Some of these things would be really, really amazing. Scary because, well, as with anything, people can abuse these sorts of models.

And you cannot control the person who wants to abuse.

Exactly. Any tool probably in human history – not any, but a lot of them can be used for good and for bad things. You have a kitchen knife that you can cut vegetables with or something…

Or cut the neighbor!

Yeah, exactly! [*both laugh*] It's the same with the technology nowadays.

Do you know what Nobel invented? Dynamite!

Yeah, Nobel, exactly. There we go! This is a very good example.

Would you agree with Yann LeCun, whom I have interviewed twice? He says that today's AI is no smarter than a home cat.

Than a cat? Well, probably not at this point, but we will see how much time will pass until they are smarter than a cat, so we'll see! [*she laughs*]

Is it exactly what you wanted to do to be at the intersection of all these nice things at the moment they are going to intercept?

Yeah, I think it's a nice place to be. It's sometimes a bit uncomfortable because things are moving really, really fast. As I said, during my PhD time, everything felt more chill, whereas nowadays, there's a lot going on, but it's also very fun.

Is this why you do research to have fun in something innovative?

I mean, partially. There are various different reasons why I do research. I think I'm bored quite easily, so I like to do things that are constantly new, so that your day to day isn't just repeating the same things over and over again. Also, because of what I just mentioned, all these amazing

things that these tools that we're developing can potentially help us unlock. But also to have fun.

I am sure you are not afraid you are going to be bored by AI in the coming years.

No, I don't think so. The opposite, actually.

In your short career until now, what in AI particularly made you say, 'wow'?

That's a very good question. I remember we had this Flamingo model in DeepMind. Before it came out, we could play around with it internally a bit, and I remember uploading a picture of my dad's cat and asking it various questions about the picture, and it could answer everything. I was like, how is this possible? That was my first moment like, wow, these things actually work! Because I was pretty much a sceptic before that about this kind of increase the model size, increase the data size. I wasn't thinking this is actually going to work. I mean, there's probably a limit. I still think there's a limit, but we're still pushing this frontier it would seem.

What is the next 'wow' you are going to say?

If we have actual embodied agents. If we can integrate all of these current assistants we have and put it into a robot This is the point where I would actually find it scary as some sort of actual sci-fi moment, but it would also be like, wow.

What would be your dream for the next 10 years?

Personally, I would like to move into the more application side of things, where you can do something good with these models. The model development is very, very interesting as well, but this has kind of been a change in the past couple of years for me, where we've moved on from models that don't really work to models that do work, and now it's unlocked this whole world of possibilities where you can actually use it to solve real-world problems.

The spectrum of things that you can do is infinite. If I only think about climate change, the number of issues that AI can help on is infinite.

Exactly.

Will you pick one of those?

Good question. I think we have so many open problems in our world, from various political problems to climate change. I think climate change is a good one, maybe something medical. I find some of the companies that do research into longevity, not necessarily extending the length of human life but improving the quality of life at a later stage where I think there's a lot of ML used there, I think these are the things that I find interesting.

You are preparing for retirement.

Yeah, exactly. I want to retire in a nice way! [*both laugh*] Still be able to go on hikes or something.

Scuba diving?

Yeah, perfect. Scuba dive at 90 or something like that! [*Ivana laughs*]

We have spoken about the future; we did not speak much about the past. Why did you leave Croatia?

Well, that was a very long time ago now. It was 12 years ago. Yeah, that's when I left Croatia. I wanted to see what there is outside Croatia. I wanted to experience life outside Croatia, but I didn't think I was going to stay. I initially thought it was going to be a couple of years, but then, here we are.

What did you do for this couple of years?

I did a Master's in Berlin at first. Then, I was in California for a year doing an internship at a startup, and then I started doing my PhD in Edinburgh, Scotland.

So, it was not the plan?

No, not initially. It was kind of let's go and see sort of thing.

Is this the way you decide things?

When I was young, for sure. Now, sometimes!

Are you going to use the same criteria for things happening in the future?

I mean, not so much. Now, I tend to overthink things a bit more than when I was 21.

What is the thing that you did until now that you are the most satisfied with?

I'm going to say something that's unrelated to my career. I ran a marathon a month and a half ago! Yeah, it was hard, but also, it made me feel like, oh, I can do this.

> ***"I think it's this thing about not giving up. Knowing that you can do more than you think you can!"***

Wow, I'm so jealous. I registered for three marathons. I was never able to start one. I always got injured.

No way. It's very annoying. I'm still a bit injured now. My feet hurt, but other than that, I enjoyed it a lot. It was one of those things because I often doubt myself, and I'm like, this is too hard, I cannot do this, I cannot do that, and it's one of those things where you're like, I can actually do this.

Four hours?

3:57.

Tell me one thing that our readers can learn from your marathon.

I think it's this thing about not giving up. Knowing that you can do more than you think you can because I feel like it's one of those things where you have to believe that you can do it. But yeah, as I said, I'm not that kind of person.

Some people will say that life is already competitive enough.

True, but I think it makes you more robust to everyday life challenges because it teaches you, yeah, I can do this. I can overcome this thing that I wasn't able to do.

I am sure that finishing a marathon makes you very, very high.

Yeah, it does. I was laughing and crying at the same time. I was in this weird state. [*Ivana laughs*]

How many CVPRs have you attended?

This is my first one. I have never been to CVPR before. I normally go to ML conferences.

How are you finding it?

Yeah, I think it's nice. It's great. I don't know how big it is, but from what I've heard, it's bigger than NeurIPS.

11,500 this year, including online.

Yeah, I'm excited for it. Let's see.

What did you learn from the workshop?

That there are a lot of different opinions and ways people do prompting. I don't know if that's a good thing or a bad thing. In language, prompting is a well-defined thing, whereas here, it's very much an open research problem open for discussion. It's possibly a good thing. There might be good things coming out of it.

Is there a lifelong dream or goal you hope to achieve before you retire?

I want to work on a particular problem that matters right now. I'm now doing research for the sake of research because it's interesting. I haven't decided which of these important problems we talked about yet is that one problem, but I want to do something like that and focus and try to help achieve that.

Tell us one thing about you that we do not know.

Again, that's a very difficult question.

I do not know much!

I don't think I like to talk about myself that much! [*she laughs*]

Oh, so you had a great idea of accepting my offer for an interview.

Yeah, to be honest, I was not sure what I was going to say!

Now, it's too late!

I know. People have convinced me otherwise.

Who convinced you?

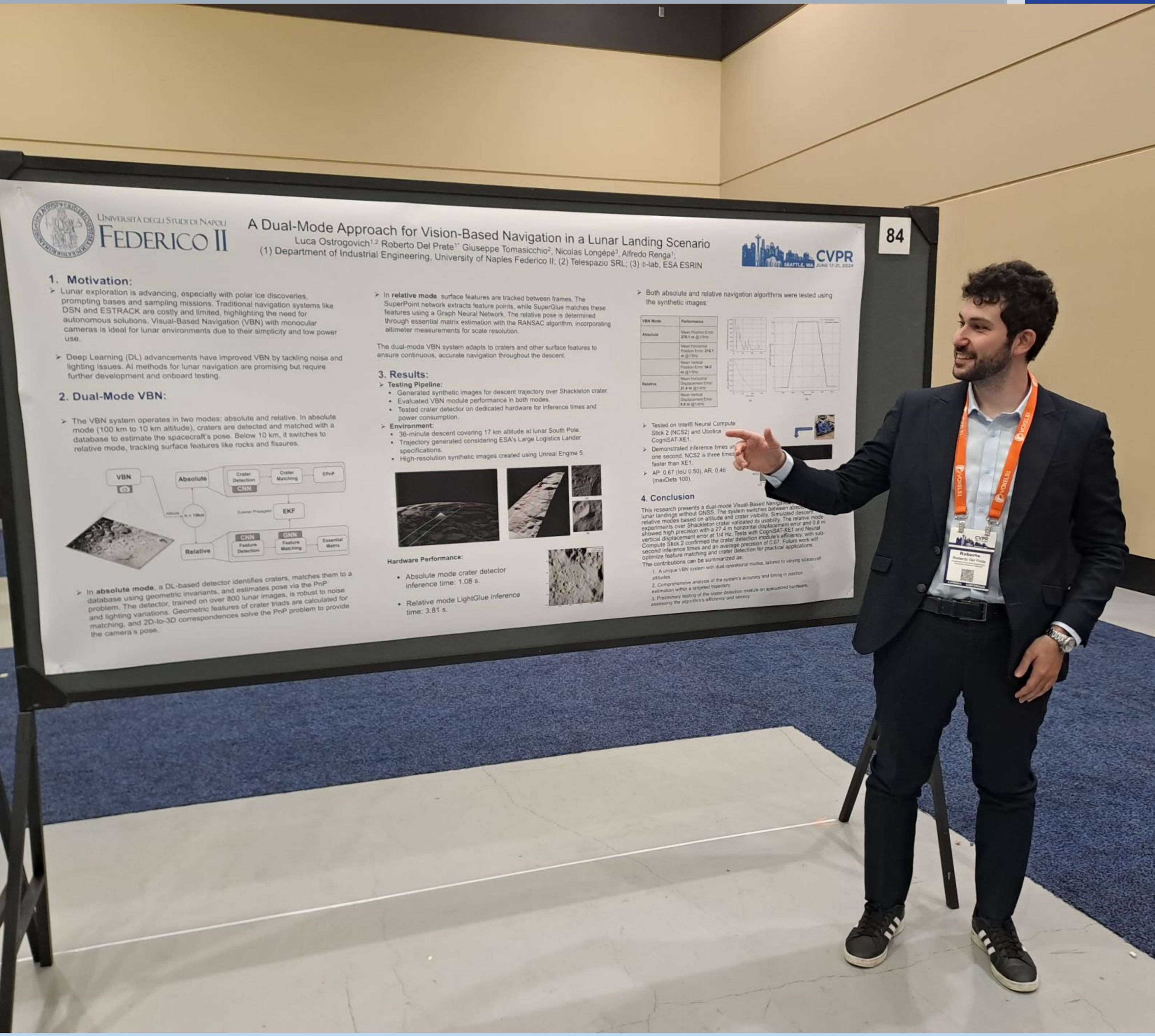
Well, my manager and my boyfriend. They were both like, *"No, you should do it. It's a good thing. It inspires people!"* I'm like: *"I don't know if I like doing these self-promotion things."*

I often discover the most surprising things about people I do not know.

Yeah, well, there we go. That's one thing you didn't know about me! [*both laugh*]

Do you have a final message for the community?

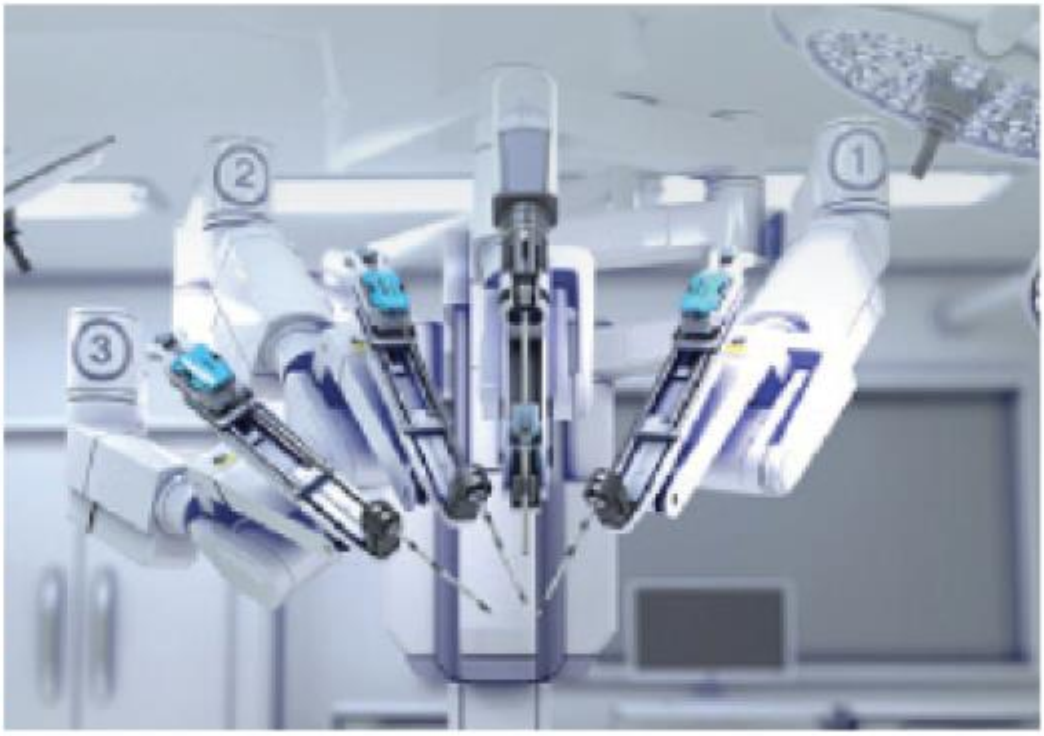
Do big things that matter because I think you can achieve it. There are so many smart people in this field and there are many, many interesting problems to be solved. I think if we put our minds together, we can improve this world that we live in.

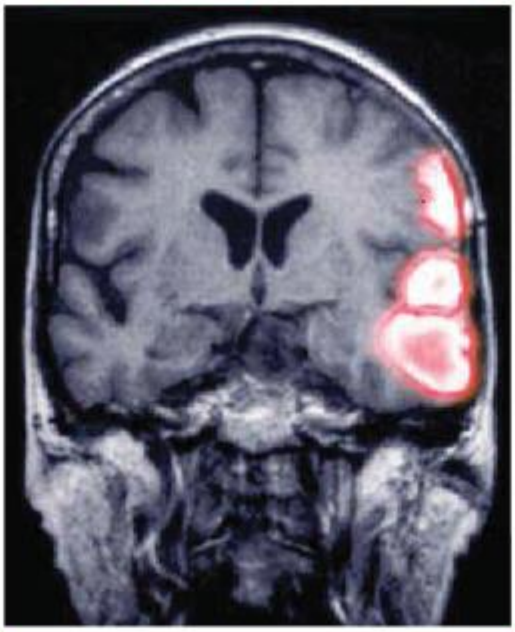

Università degli Studi di Napoli Federico II
A Dual-Mode Approach for Vision-Based Navigation in a Lunar Landing Scenario
Luca Ostrogovich1,2 Roberto Del Prete1* Giuseppe Tomasicchio2, Nicolas Longépé3, Alfredo Renga1;
(1) Department of Industrial Engineering, University of Naples Federico II; (2) Telespazio SRL; (3) Φ-lab, ESA ESRIN
CVPR Seattle, WA June 17-21, 2024
84
1. Motivation:
2. Dual-Mode VBN:
3. Results:
4. Conclusion
Hardware Performance:
VBN
Absolute
Relative
EKF

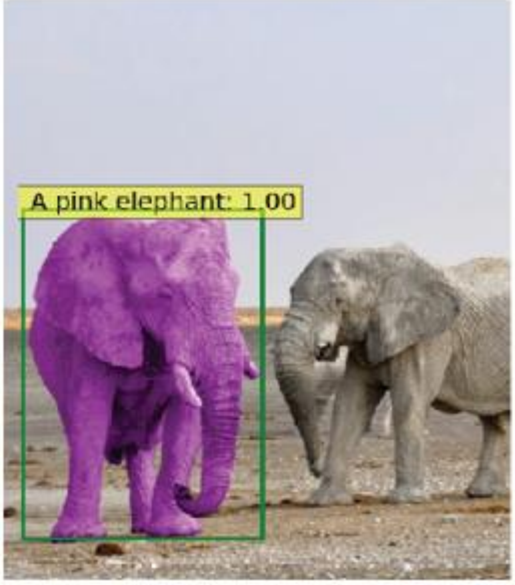
Roberto Del Prete, a PhD student from the University of Napoli, who is set to complete his PhD in November, is proud to present at his first CVPR conference.

His poster, presented at the AI4Space workshop, discusses a novel approach for autonomous lunar landing, leveraging visual information.

The workshop ighlighted the space capabilities that draw from and/or overlap significantly with vision and learning research, outline the unique difficulties presented by space applications to vision and learning, and discuss recent advances towards overcoming those obstacles.






A pink elephant: 1.00