1

Why Gait Needs a Special Backbone

A silhouette isn't a selfie

CORAL BITES
SEAFOOD & SHAKES

REEF BOULEVARD

Wait... why do CNNs keep slipping on these silhouettes?

OPEN

TODAY'S SPECIAL: SEA SALAD

1

2

Because gait is sparse, binary, and full of tiny motion clues!

3 TRADITIONAL CNN BACKBONE

SLIP!

Too flat! No grip!

4

WHY IT'S HARD

=

Most pixels are zero. Only a few lit pixels carry the motion clues!

5

GAIT TRANSFORMER BACKBONE

- ✔ END-TO-END
- ✔ GLOBAL ATTENTION
- ✔ LONG-RANGE LINKS
- ✔ CAPTURES TINY MOTION CLUES!

SMART ATTENTION!

MOTION CLUES CAPTURED!

6

VIP

IDENTIFIED

POP!

Transformer nailed it!

That's our special backbone!

MORE STEPS.
MORE CLUES.
MORE CONFIDENCE.
RECOGNIZED!

Gait Transformer: See the Walk, Know the Who!

PAGE 2

The Gap in Prior Transformer Gaits

Half a solution is still wobbly

Gait Frames (Sequence)

Patch Tokens

GaitViT used a vanilla ViT on GEI... but temporal modeling was weak.

GEI (Averaged Silhouette)

Time

1 2 3 ··· T

Weak!

SwinGait helped, but leaned on shallow CNNs and bridges.

SwinGait (Windowed + Shallow CNNs)

Windows

Shallow CNN

Shallow CNN

Shallow CNN

Bridges

GRAAAAH!

CORAL GRILL

BEST KRABBY PATTIES!

The fix? An end-to-end Transformer backbone that unites all frames!

End-to-End Transformer Backbone

Gait Recognition

Global Attention Across Time

STRONG TEMPORAL MODELING!

End-to-end Transformers seal the gap in temporal modeling—no more wobbles!

POP!

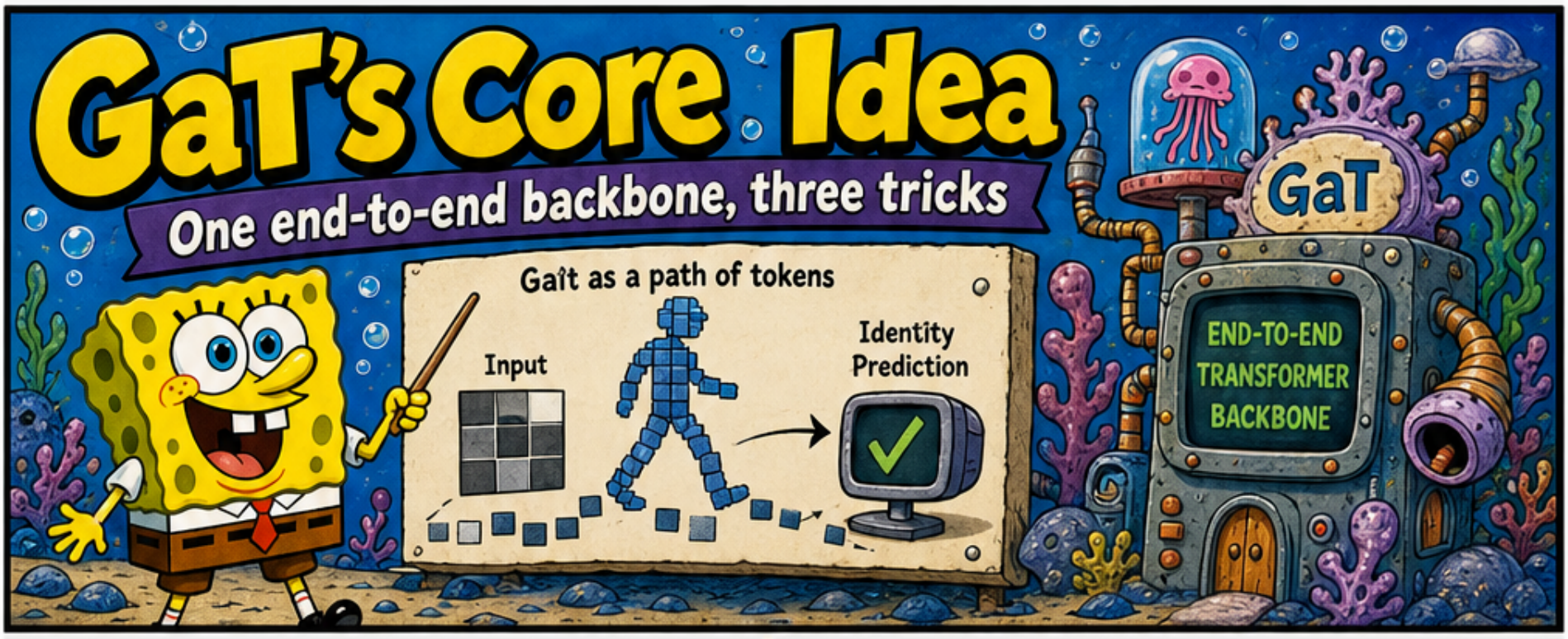
FULL MODEL. FULL CONTEXT.

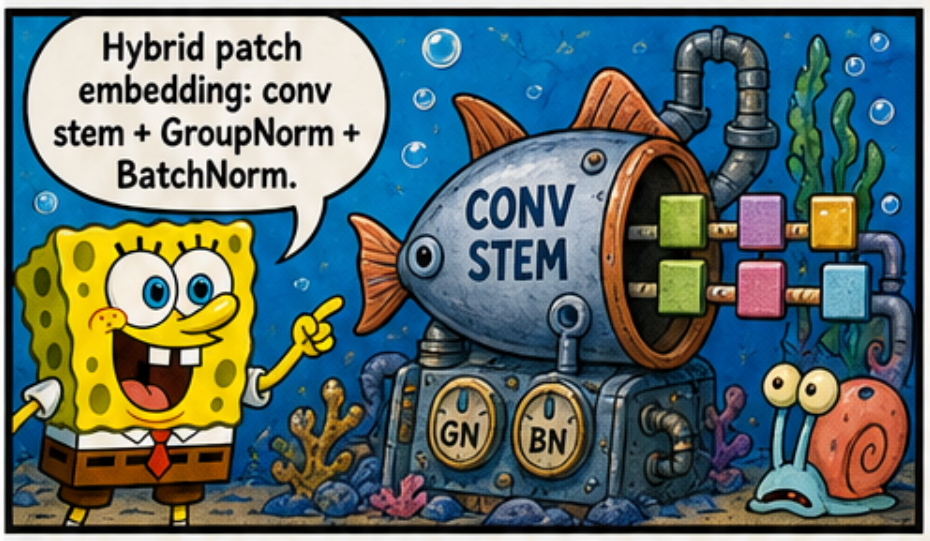
FULL CONFIDENCE!

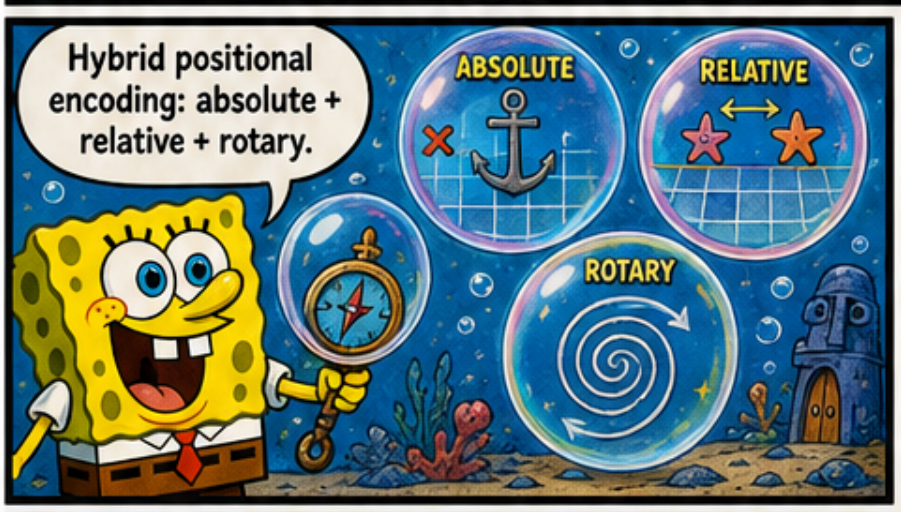
Better modeling. Better recognition. Better adventures!

GaT's Core Idea

One end-to-end backbone, three tricks







4

How It Helps Under Data Scarcity

Built for small, noisy gait datasets

REEFBAY RESEARCH LAB

Today I'll show how our Gait Transformer works—even with tiny, noisy data!

END-TO-END PIPELINE

RAW GAIT SEQUENCES → TOKEN EMBEDDING → TRANSFORMER BACKBONE → IDENTITY (WHO?)

Sounds fancy. But we barely have data!

CORAL TOWN

PROBLEM: DATA SCARCITY & NOISE!

Just a handful of messy walks!

OUR DATASET SO SMALL!

NOISY! BLURRY! UNDERWATER!

STEP 1: TURN NOISY WALKS INTO CLEAN TOKENS!

NOISY FRAMES (VERY FEW)

CLEAN TOKENS

TOKENIZER 3000

I squeeze out the noise and make neat tokens!

STEP 2: STRONG INDUCTIVE PRIORS INSIDE THE TRANSFORMER!

Strong inductive priors help when data is limited!

TRANSFORMER BACKBONE

ATTENTION LOOKS ACROSS TIME!

LEARN PATTERNS THAT GENERALIZE!

Underwater currents? No problem!

Different views? Bring it on!

STEP 3: TRAIN FROM SCRATCH (NO PRETRAINING!!

No pretraining needed—train from scratch.

TRAINING... 100%

LEARNING FAST!

SMALL BUT MIGHTY DATA

Whoa, that was quick!

RESULT: IT WORKS WITH TINY, NOISY DATA!

INPUT: FEW & NOISY

PREDICTED IDENTITY

BIGFIN

CONFIDENCE: 98.7%

Even with so little data, I've got it!

SURPRISE TEST: WILL IT RECOGNIZE A LEGEND?

?!

IT'S BIGFIN! I KNEW IT!

RECOGNIZED!

IDENTITY: BIGFIN

CONFIDENCE: 99.3%

THE END!

Takeaway: Better Gait Features, Cleaner Design

One backbone to rule the reef

THE KRUSTY KRAB

FRESH KRABBY FEATURES!

Let's build the ultimate gait model!

1) INPUT: WALKING OVER TIME

t_1 t_2 t_3 ... t_T

SPATIAL TOKENS (Per frame)

Gait Transformer Backbone

TEMPORAL TOKENS (Over time)

CLEAN • UNIFIED • POWERFUL
One backbone to rule the reef!

SPATIAL: How you move (at each moment)

TEMPORAL: How you evolve over time

GaT jointly captures spatial and temporal variation.

It achieves state-of-the-art on Gait3D, GREW, and CCGR-MINI.

WAY TO GO!

NEW SOTA!

GAIT3D	GREW	CCGR-MINI
Accuracy	Rank-1	mAP
99.2%	95.7%	91.8%

BUT THEN...

A NEW CHALLENGER?!

ORDER UP!

HARD GALI DATA

DOMAIN SHIFT SPECIAL

You thought it'd be that easy?

HA HA HA!