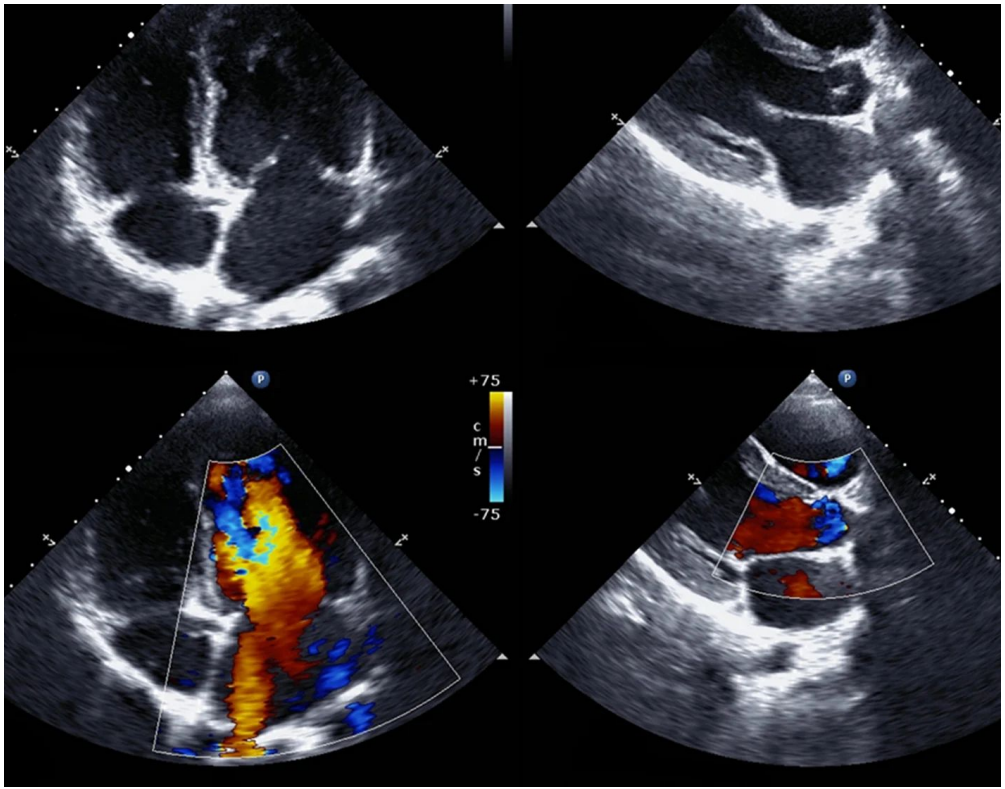
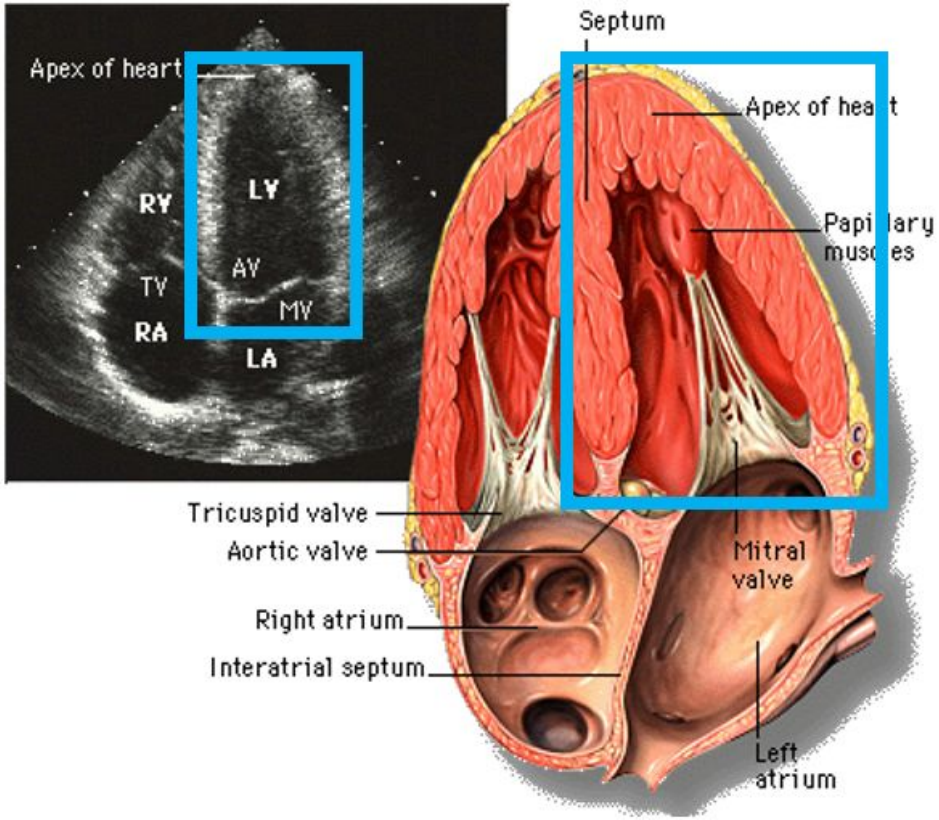
Auto-doctor of echocardiography

Chengche Tsai, Alessandro Folloni

Echocardiography



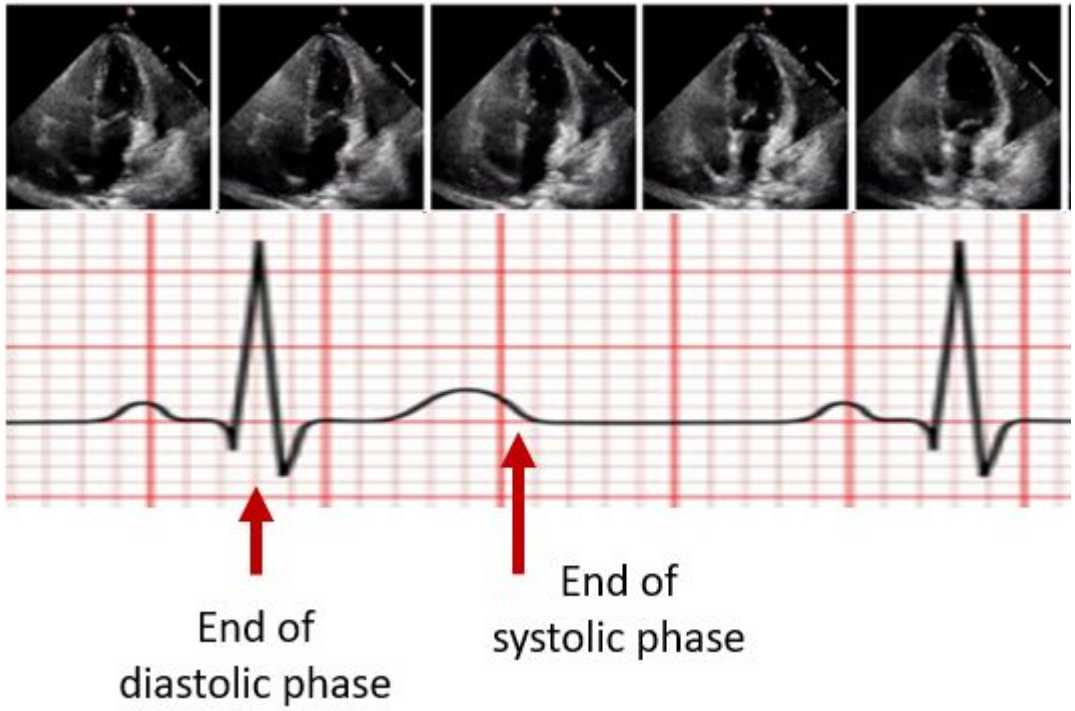
What matters most in heart echo?

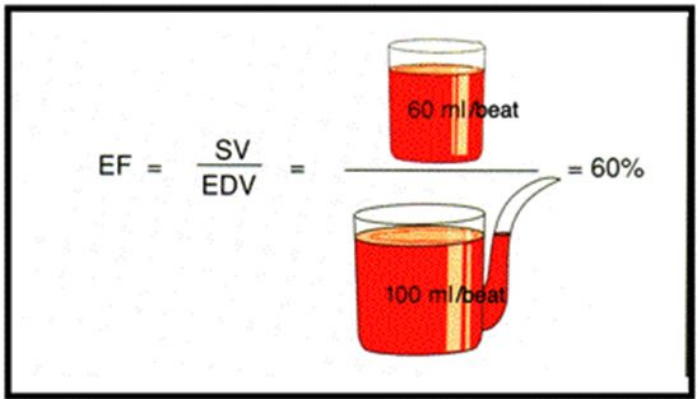


Left ventricle:

- Segmentation
- Ejection fraction estimation

Ejection fraction (how much blood the heart pumps out)





Past works - Image-based approaches

Three-stage approach:

- Stage 1: localize the frames that belong to ED and ES.
- Stage 2: train/apply a model doing segmentation on these 2 frames.
- Stage 3: Apply math formulas with a bunch of geometric assumptions to the segmentation masks and acquire the volumes.

Downside:

- **No spatiotemporal understanding.**
- **Lost information between ED and ES frames.**

Past works - video-based approaches

- Single-task models
- **Heavily rely on the position pointers to know where ED/ES are.**
- Only **three papers** have completed on large-scale datasets.
- The performance
  - Segmentation dice is good (>0.9)
  - EF estimation has room to improve

Past works - video-based approach

	Need ED/ES location?	MAE	RMSE	R^2
EchoNet (AF)	x	7.35	N/A	0.4
EchoNet (AF)	v	4.05	5.32	0.81
Video transformer	v	5.32	7.23	0.64
R3D	v	4.22	5.62	0.79
EchoGraphs	v	4.01	5.36	0.81
EchoCoTr	v	3.95	5.17	0.82
EchoGNN	x	4.45	N/A	0.76

Our goals

A video-based model that

1. Does not rely on ED/ES annotations
2. Can provide left ventricle segmentation masks
3. Can estimate ejection fractions

Models to explore

- TimeSformer
- Models with 3D reconstruction / 2D projection
- Segment anything (SAM) and its medical variants
- R-CNN

Datasets for our task

- CAMUS (2019), France
- EchoNet Dynamic (2021), Stanford

CAMUS (2019)

Data recruitment:

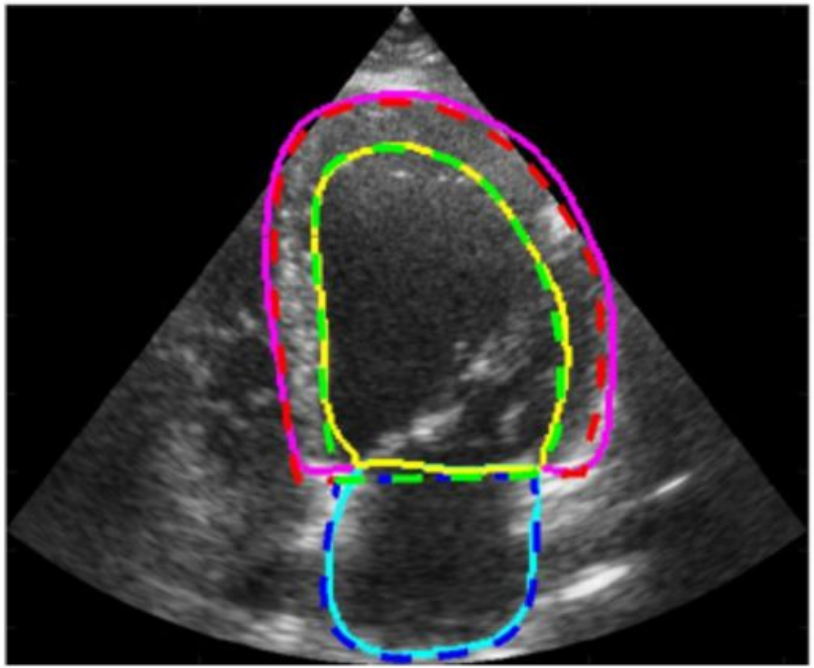
- Videos from 500 patients, acquired at the University Hospital of St Etienne (France).
- In order to enforce clinical realism, neither prerequisite nor data selection have been performed.
- Availability of paired ECG is questionable.

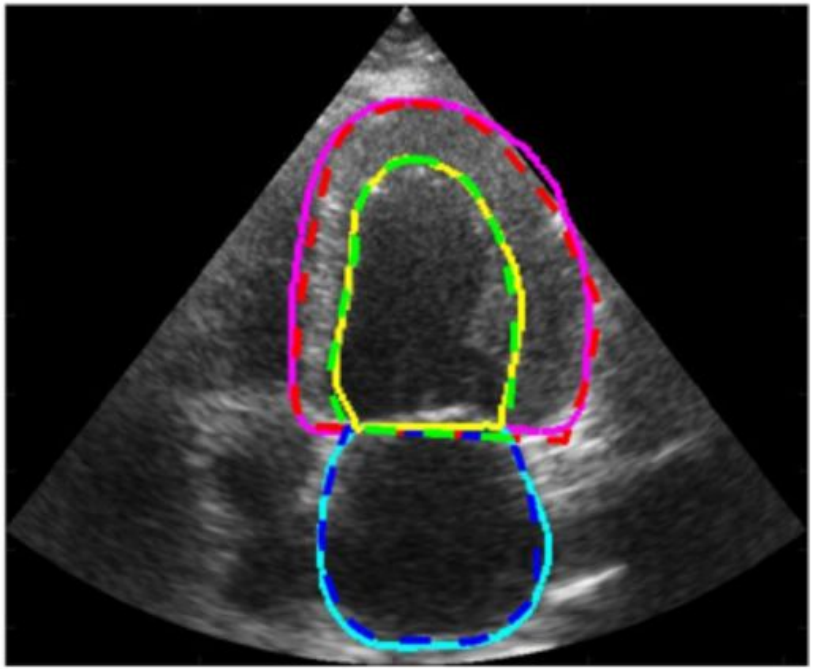
Data description:

- No fixed frames videos.
- Full anatomical annotations for the left ventricle; ED and ES annotated.
- Each video has a different resolution, and all of them are larger than 1024 × 512.

Uniqueness: One subset was annotated by the same physician 7 months apart
→ to measure intra- and inter-operator variability.

CAMUS - example





EchoNet Dynamic (2021)

Data recruitment:

- A standard full resting echocardiogram study consists of 10,030 apical-4-chamber echocardiography videos.
- It has paired ECG

Data description:

- Each video was cropped and masked to remove medical text and information outside of the scanning sector.
- The resulting images were downsampled into standardized 112x112 pixel videos.
- Every video contains one or more cardiac cycles (20 - 40 frames per video).

EchoNet Dynamic - example