**Slide 1-2:**

- Em xin kính chào thầy cô và các quý doanh nghiệp, hôm nay nhóm nghiên cứu gồm chị Nhâm Đỗ Hải Ninh và em Nguyễn Việt Dũng sẽ trình bày về các kết quả của nghiên cứu Phân đoạn và xác định vị trí cá trong môi trường nước.

- Báo cáo của chúng em bao gồm ba phần…

**Slide 3-4:**

- Ngành nuôi trồng thủy sản phụ thuộc lớn vào việc đưa ra ước lượng về kích thước, trọng lượng và xác định vị trí của cá để đưa ra các dự đoán về cân nặng, giới tính và độ tăng trưởng của cá. Tuy nhiên, các phương pháp truyền thống vô cùng mất thời gian và công sức.

- Hơn nữa, kể cả với mô hình phân vùng ảnh có thể tự động hóa việc đo đạc thì việc gán nhãn cho các điểm dữ liệu đòi hỏi sự chính xác và tỉ mỉ, cần nhân lực, thời gian và chi phí.

- Mặt khác, kiểm soát môi trường sinh sống của cá là một công việc quan trọng trong ngư nghiệp giúp đánh giá tình trạng sức khỏe, chế độ dinh dưỡng, có nên đánh bắt hay không.

- Nếu những loài cá nằm trong danh sách cần bảo vệ do có nguy cơ tuyệt chủng bị đánh bắt nhầm có thể dẫn đến mất cân bằng sinh học. Do đó, việc nắm bắt được thông số của cá có ứng dụng thực tiễn ko nhỏ.

Bài nghiên cứu của chúng em đề xuất mô hình học sâu với nhiệm vụ khắc phục các hạn chế trên.

**Slide 5-6:**

- S5: Mô hình của chúng em nhận input là ảnh cá, đầu ra là ảnh phân vùng cá. Nhãn của dữ liệu huấn luyện có hai loại bao gồm nhãn đầy đủ và nhãn điểm.

*Trên hình vẽ (chỉ tay) là ảnh input, ảnh groundtruth mask (có 2 ảnh đấy) và ảnh được gán nhãn điểm, ảnh phân vùng (nhớ phải chỉ tay).*

- S6: Như vậy, chúng em sẽ thực hiện 3 nhiệm vụ, phân đoạn dựa trên giám sát yếu (weakly-supervised segmentation); xác định vị trí cá (localization) và đếm cá (counting)

**Slide 7:** Tiếp theo chúng em sẽ giới thiệu mô hình đề xuất

**Slide 8-12:**

- S8: Tích chập thông thường (bên trên) và tích chập atrous (bên dưới). Ta thấy đối với tích chập atrous thì các receptive field giãn cách nhau một cell mặc dù về bản chất vẫn là tích chập với bộ lọc 3x3 nhưng được thực hiện trên một vùng rộng hơn là 5x5. Tích chập thông thường thì kích thước receptive field bằng với kích thước bộ lọc và bằng 3x3. Do đó ASPP giúp nhận diện các vật thể có thể xuất hiện trong ảnh với nhiều kích thước khác nhau.

- S9: Tuy nhiên, đối với mô hình đề xuất chúng em sử dụng một phiên bản mạnh hơn của ASPP, đầu vào của module là 1 feature map với C chiều, feature map này được tách thành 4 phần và được đưa vào 4 khối tích chập song song ta thu được 4 feature maps . Mỗi feature map thu được chỉ có kích thước chiều sâu bằng 1/4 chiều sâu của feature đầu vào, đọc tiếp cho đến biểu thức 1. (không đọc conv3)

- S10: Cộng 2 nhánh chứa lớp tích chập atrous liền nhau rồi tiếp tục cộng dồn với feature map ở 2 nhánh residual, ta nhận được (như trên slide)

- S11: Các $F\_{o^{'}\_i}$ cần được kết hợp với nhau để tạo nên feature output đủ mạnh, ta có biểu thức concatenate như trên slide…

Ở đây $F\_{o^{''}\_1}$ khử feature với receptive field nhỏ, $F\_{o^{''}\_2}$ làm ngược lại; cả 2 đều có chiều sâu bằng 1 nửa chiều sâu của input feature. Bước cuối cùng khi tất cả các feature được ghép lại với nhau.

**Slide 13:** Baseline PASPP-FCN8-wide-ResNet50-2: Baseline chúng em đề xuất là …; bao gồm 1 mạng FCN8 được pretrained trên wide-resnet50-2 (đọc wide-resnet-50); ở giữa encoder và decoder là module PASPP như trên slide… 5 layer bao gồm khối màu tím và layer 1 -4 là pretrained của wide-resnet-50.

**Slide 14-19:** Trước khi đi vào mô hình đề xuất, em sẽ trình bày về affinity matrix hay ma trận tương đồng

- S16: Đây là một công cụ thống kê có vai trò biểu diễn sự tương đồng, “thân cận” của các điểm dữ liệu. Trong nghiên cứu của Ahn và Kwak, mạng AffinityNet được sử dụng để dự đoán semantic affinity giữa 2 pixel trên 1 ảnh huấn luyện. Ma trận affinity dự đoán được sẽ được biến đổi thành ma trận xác suất chuyển và đưa vào module Random Walk để có thể mở rộng (propagate) vùng activation scores của activation map đến các vùng xung quanh mà có mức tương đồng cao, từ đó chất lượng phân vùng của activation map được tăng lên.

- S17: Đối với đồ thị vô hướng, ma trận affinity là một ma trận vuông cấp n với n là số đỉnh của đồ thị, mỗi giá trị ở hàng i cột j tượng trưng cho độ tương đồng giữa node/đỉnh i và node/đỉnh j. Đối với các bài toán trong lĩnh vực thị giác máy tính, ta tưởng tượng mỗi bức ảnh là một đồ thị có trọng số với mỗi pixel là 1 đỉnh và trọng số trên mỗi cạnh biểu thị sự tương đồng giữa mỗi pixel tùy theo đặc thù của bài toán, các trọng số này sẽ hình thành một ma trận affinity với số cấp là số pixel của bức ảnh. Đối với image segmentation, ma trận affinity nên cho thấy sự tương đồng giữa các đặc tính (features similarities) giữa mỗi pixel. Để thu được ma trận này, một ảnh sau khi được trích chọn đặc trưng và cho ra một feature map, ma trận affinity W thường được tính toán như trên slide…

- S19: Sau khi có được ma trận affinity $W$ ta xây dựng ma trận xác suất chuyển theo công thức…

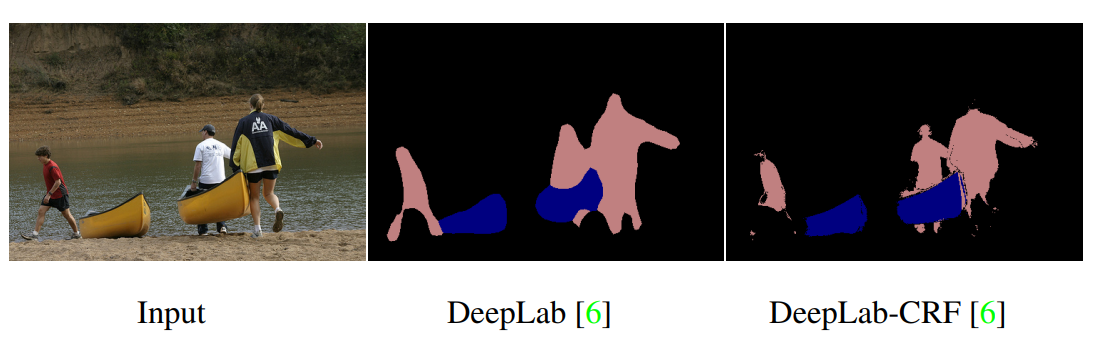
Siêu tham số $\beta > 1$ giúp mô hình bỏ qua các features có độ tương đồng thấp.

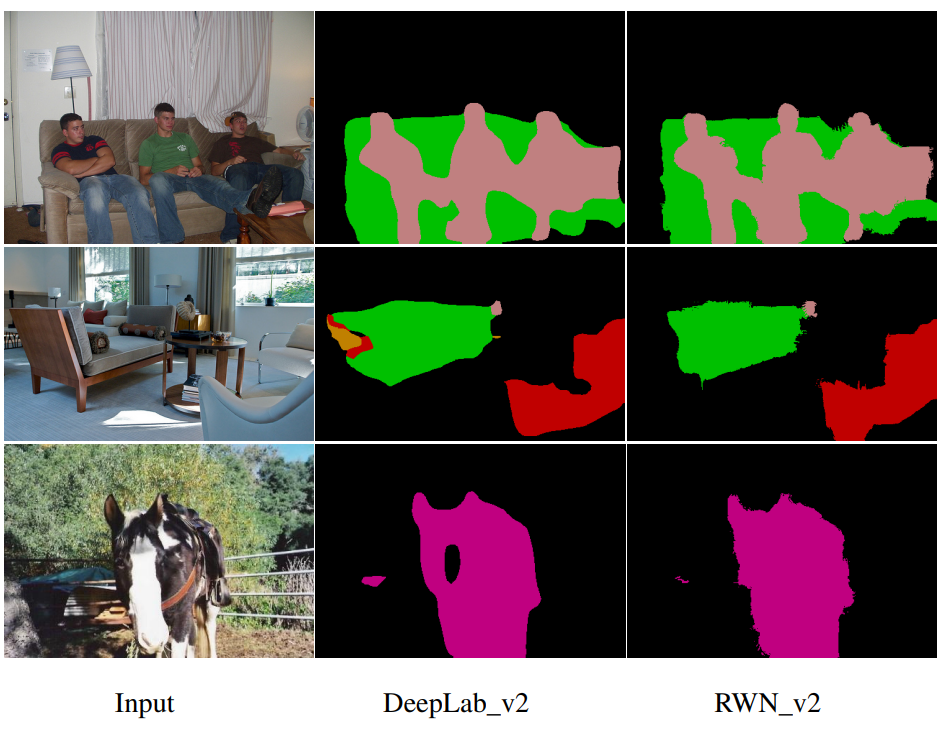
Giải thuật random walk được lặp hữu hạn lần với score ban đầu $vec(M\_c)$

\* không nói chỗ này, có hỏi thì trả lời:

(Nếu không dùng Random Walk thì không thể phân tán các activation scores dc)

(Affinity thay thế cho CRF trong nghiên cứu của Ahn và Kwak giúp giảm khối lượng tính toán để đưa ra ma trận xác suất chuyển)





- S19: (đọc slide)

**Slide 20-25:** LCFCN

- S21: đọc đoạn đầu, bỏ công thức và bỏ chỗ giải thích mấy cái L

- S22: đọc như trên slide trừ công thức

- S23: đọc như trên slide trừ công thức

- S24: đọc 2 dòng đầu trước công thức; sau đó phải nói đại loại 1 câu là, xét pixel k được gán nhãn điểm có thuộc background hay k

- S25: chỉ nói qua về cơ chế split, k đi sâu vào 2 pp

- S26: như trên slide trừ công thức

**Slide 28-34:**

- S28: (slide)

- S29: Mỗi bộ dữ liệu con bao gồm 50\% training, 20\% validation và 30\% testing; và đảm bảo cân bằng lớp giữa background và foreground. Tập \textit{FishLoc} có 1600 ảnh training, 640 ảnh validation, 960 ảnh testing và 310 ảnh training, 124 ảnh validation, 186 ảnh testing cho tập \textit{FishSeg}.

- S30: (IoU) Là chỉ số đánh giá được sử dụng để đo độ chính xác của object detector trên tập dữ liệu cụ thể; là kết quả của diện tích phần giao chia cho diện tích phần hợp.

- S31: (Accuracy) đánh giá tỷ lệ mà mô hình phân chia chính xác các pixel vào các lớp. Được tính bằng tỉ lệ dự đoán đúng chia cho tổng số dự đoán.

- S32: (MAE) …với $C\_i$ là số lượng cá thực tế và $\hat{C\_i}$ là số lượng cá đếm được trong ảnh thứ $i$.

- S33: \textbf{GAME.} Việc xác định vị trí của cá khó hơn so với việc phân loại và đếm vì trong môi trường nước nếu số lượng cá nhiều có thể dẫn đến nhiều vị trí trùng nhau. MAE không thể thể hiện rõ độ chính xác trong việc xác định vị trí vì mô hình có thể đếm nhầm cá mà vẫn cho kết quả đáng tin cậy. Do đó cần 1 metric xét cả 2 việc đếm và xác định vị trí của cá; do đó sử dụng Grid Average Mean absolute Error (GAME)...

- S34: (trên slide)

**Slide 35-47:**

- S35: (trên slide)

- S36-37: pretrained, PASPP

- S38-39: pretrained, PASPP

- S40-41: trên slide

- S42: (bổ sung)

- S43-45: (trên slide)

3/ So sánh adjacency vs affinity matrix? Tại sao k dùng adjacency matrix mà dùng affinity matrix?

Adjacency chỉ gồm 1 và 0, affinity matrix chứa hệ số tương quan sữa các pixels. Dùng affinity matrix có tác dụng phân tán confidence score dựa trên mức độ tương quan.

4/ Tại sao dùng chuẩn 1 mà k dùng chuẩn 2, như gauss? (pre-ans: tính toán đơn giản hơn, đỡ oom, như kiểu bộ lọc, ngta hay dùng) Tại sao dùng e mũ? Tại sao có dấu - trong ngoặc?

Chuẩn 1 tính toán đơn giản hơn. Một nghiên cứu về phương pháp phân cụm phổ (spectral clustering) cho thấy dùng affinity dạng e mũ có tác dụng phân tách hiệu quả cao hơn. Dấu trừ khiến cho các pixels có feature vector cách xa nhau thì độ tương quan thấp đi.

5/ Không đánh weight cho CAM ~ Tại sao dùng hàm focal?

**Slide 14: Du động ngẫu nhiên trong phân vùng ảnh**

Sau khi activation map được huấn luyện qua AffinityNet, affinities được xét trong 1 vùng bán kính \gamma, nhận được ma trận affinity W có phần tử 1 nằm trên đường chéo chính, ta tính toán ma trận xác suất chuyển như trên slide:...

Ở đây siêu tham số \beta>1 nên lũy thừa Hadamard W^{o\beta} của ma trận affinity ban đầu bỏ qua những affinities không quá quan trọng trong W. Dùng W^{o\beta} thay vì W giúp cho mask tạo được gần với thực tế hơn.

Sau khi thuật toán Random Walk được thực hiện, tức là T được nhân với CAM sau 1 số vòng lặp sẽ nhận được CAM chất lượng cao M\_c^\* của lớp c với như trên slide…

với vec(.) là vector hóa của ma trận, t là số vòng lặp.

**Slide 15-20: Phương pháp LCFCN**

- S15: đọc đoạn đầu, bỏ công thức và bỏ chỗ giải thích mấy cái L

- S16: đọc như trên slide trừ công thức

- S17: đọc như trên slide trừ công thức

- S18: đọc 2 dòng đầu trước công thức; sau đó phải nói đại loại 1 câu là, xét pixel k được gán nhãn điểm có thuộc background hay k

- S19: chỉ nói qua về cơ chế split, k đi sâu vào 2 pp

- S20: như trên slide trừ công thức

**Slide 21-28: ASPP**

- S21: Công thức trên slide; với mỗi một cell i trên output y, tích chập atrous sẽ tính toán bằng cách nhân tích chập bộ lọc w với feature map x.

- S22: Tích chập thông thường (bên trên) và tích chập atrous (bên dưới). Ta thấy đối với tích chập atrous thì các vị trí của receptive field giãn cách nhau một cell mặc dù về bản chất vẫn là tích chập với bộ lọc 3x3 nhưng được thực hiện trên một vùng rộng hơn là 5x5. Tích chập thông thường thì kích thước receptive field bằng với kích thước bộ lọc và bằng 3x3.

Giúp chúng ta có thể nhận diện các vật thể có thể xuất hiện trong ảnh với nhiều kích thước khác nhau.

- S23-28: để ý nói tùy ý

- S8: Tích chập thông thường (bên trên) và tích chập atrous (bên dưới). Ta thấy đối với tích chập atrous thì các vị trí của receptive field giãn cách nhau một cell mặc dù về bản chất vẫn là tích chập với bộ lọc 3x3 nhưng được thực hiện trên một vùng rộng hơn là 5x5. Tích chập thông thường thì kích thước receptive field bằng với kích thước bộ lọc và bằng 3x3. Do đó ASPP giúp nhận diện các vật thể có thể xuất hiện trong ảnh với nhiều kích thước khác nhau.

- S9: Ký hiệu feature map đầu vào của module là $F\_{in}$, từ 4 khối tích chập song song ta thu được 4 feature maps $F\_{conv\_1}, F\_{conv\_2}, F\_{conv\_3}, F\_{conv\_4}$. Mỗi feature map thu được chỉ có kích thước chiều sâu bằng 1/4 chiều sâu của feature đầu vào

- S10: Cộng 2 nhánh chứa lớp tích chập atrous liền nhau rồi tiếp tục cộng dồn feature map ở 2 nhánh residual, ta nhận được (như trên slide)

- S11: với $F\_{o^{'}\_i}$ là output nhánh $i$. $F\_{o^{'}\_i}$ được kết hợp với nhau để tạo nên feature output đủ mạnh, ta có biểu thức concatenate như trên slide…

Ở đây $F\_{o^{''}\_1}$ khử feature với receptive field nhỏ, $F\_{o^{''}\_2}$ làm ngược lại; cả 2 đều có chiều sâu bằng 1 nửa chiều sâu của input feature. Bước cuối cùng khi tất cả các feature được ghép lại với nhau.

- S12: (hình vẽ)

**Slide 29-33:**

- S29-31: skip

- S32-33: (như trên slide)

Slide 34-56:

- S35:

Gán nhãn điểm (Point-level annotations): Các tác giả bài báo đã thêm gán nhãn điểm bằng cách nhấp chuột vào từng con cá. Mục tiêu của gán nhãn điểm là cho phép các mô hình học cách đếm cá

Chú thích pixel (Per-pixel annotations): Mục tiêu là huấn luyện mô hình phân đoạn cá. Kết quả của phân đoạn có thể được sử dụng để ước tính kích thước, hình dạng và trọng lượng của cá.

- S36: (trên hình vẽ)

- S37: Các tác giả chia tập DeepFish thành các tập con phục vụ cho từng task riêng biệt: FishClf để phân loại, FishLoc để đếm và xác định vị trí, FishSeg cho phân đoạn. Mỗi bộ dữ liệu con bao gồm 50\% training, 20\% validation và 30\% testing; và đảm bảo cân bằng lớp giữa background và foreground.

- S38: (như slide)

\begin{frame}{DeepFish dataset}

Trong bài nghiên cứu này, chúng em thực hiện 3 nhiệm vụ: phân đoạn dựa trên giám sát yếu và giám sát đầy đủ (fully- and weakly-supervised segmentation) trên tập \textit{FishLoc, FishSeg}; xác định vị trí cá (localization) trên tập \textit{FishLoc}; đếm cá (counting) trên tập \textit{FishLoc}. Tất cả 3 nhiệm vụ được thực hiện với các techniques khác nhau nhưng cùng 1 baseline trước khi kết quả thực nghiệm chứng minh baseline và hàm mất mát đề xuất cho kết quả vượt trội các phương pháp khác.

\end{frame}