Beyond the DVH: Voxel-based Automated Planning and Quality Assurance

Presented by
Chris McIntosh
Research Associate
Department of Medical Imaging & Physics
Princess Margaret Cancer Centre
Beyond the DVH

Meets the DVO
Fails visual inspection
Digit Recognition

- Computers can now recognize digits and faces with an accuracy that surpasses the average person.
  - Advances in both the mathematics and cloud computing

http://spie.org/newsroom/facial-recognition

DeepFace: Closing the Gap to Human-Level Performance in Face Verification

Yaniv Taigman  Ming Yang  Marc’Aurelio Ranzato  Lior Wolf
Facebook AI Group  Mentu Park, CA, USA
{yaniv, mingyang, ranzato}@fb.com
Tel Aviv University  Tel Aviv, Israel
wolf@cs.tau.ac.il

Abstract

In modern face recognition, the conventional pipeline consists of four stages: detect = align = represent = classify. We revisit both the alignment step and the representation step by employing explicit 3D face modeling in order to apply a piecewise affine transformation, and derive a face representation from a nine-layer deep neural network. This deep network involves more than 120 million parameters using several locally connected layers without weight sharing, rather than the standard convolutional layers. Thus we trained it on the largest facial dataset to-date, an identity labeled dataset of four million facial images belonging to more than 4,000 identities. The learned representations coupling the accurate model-based alignment with the large facial database generalize remarkably well to faces in unconstrained environments, even with a simple classifier. Our method reaches an accuracy of 97.25% on the Labeled Faces in the Wild (LFW) dataset, reducing the error of the current state of the art by more than 25%, closely approaching human-level performance.

toward tens of thousands of appearance features in other recent systems [5, 7, 2].

The proposed system differs from the majority of contributions in the field in that it uses the deep learning (DL) framework [3, 21] in lieu of well engineered features. DL is especially suitable for dealing with large training sets with many recent successes in diverse domains such as vision, speech and language modeling. Specifically with faces, the success of the learned net in capturing facial appearance in a robust manner is highly dependent on a very rapid 3D alignment step. The network architecture is based on the assumption that once the alignment is completed, the location of each facial region is fixed at the pixel level. It is therefore possible to learn from the raw pixel RGB values, without any need to apply several layers of convolutions as is done in many other networks [19, 21].

In summary, we make the following contributions: (i) The development of an effective deep neural net (DNN) architecture and learning method that leverage a very large labeled dataset of faces in order to obtain a face representation that generalizes well to other datasets; (ii) An effective facial alignment system based on explicit modeling of 3D
Learning in RT

• Recognize a patient, not a digit
• Treat you like we treated your “twin”
• We wish to derive or “learn” relationships:
  – Patient – Patient
  – Plan – Plan
  – Plan - Patient
• Can use these relationships to infer knowledge:
  – Patient → Plan
  – Plan → Complexity
  – Patient, Plan → Quality
The “Black Box”

Computer

Zero
A Patient Vector

• Typically patient features are concatenated into a vector
  – e.g. Raw image or image features
• Vector is used to quantify the patient
• Vector becomes the input the machine learning system
• The system is trained to predict a desired output
• Learn how our experts visually assess a patient image and plan

![Neural network diagram]
Deep Learning
Patient Similarity

• What does similar mean between patients?
Plan Feature Extraction

- Extract features:
  - Control point geometry and weighting
  - Fraction weighting
  - ROI features from classed ROIs
  - Dose map shape at varying isodose contours
## ROI Shape & Appearance

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometrical Centroid</td>
<td>The centre of the ROI</td>
</tr>
<tr>
<td>Principle Component Analysis -</td>
<td>Variance along the three primary modes of variation</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
</tr>
<tr>
<td>Principle Components</td>
<td>Primary modes of variation</td>
</tr>
<tr>
<td>Volume</td>
<td>Volume of the ROI</td>
</tr>
<tr>
<td>SDF Histogram</td>
<td>Histogram of the shape of the ROI</td>
</tr>
<tr>
<td>Maximum Thickness</td>
<td>Thickness of the ROI computed from the SDF</td>
</tr>
<tr>
<td>Mean Intensity</td>
<td>Mean image intensity within the boundaries of the ROI</td>
</tr>
<tr>
<td>Mode Intensity</td>
<td>Mode image intensity within the boundaries of the ROI</td>
</tr>
<tr>
<td>Standard Deviation of Intensity</td>
<td>Standard deviation of image intensity within the boundaries of the ROI</td>
</tr>
<tr>
<td>Minimum Intensity</td>
<td>Minimum image intensity within the boundaries of the ROI</td>
</tr>
<tr>
<td>Maximum Intensity</td>
<td>Maximum image intensity within the boundaries of the ROI</td>
</tr>
<tr>
<td>Intensity Histogram</td>
<td>Histogram of intensity values within the boundaries of the ROI</td>
</tr>
</tbody>
</table>

[Images (a), (b), (c)]
Similar patients receive similar treatment at the voxel level.

How similar are two patients?
- Not just intensity
- ROI shape and appearance (Radiomics)
- OAR and target geometry
  - Target near heart, etc.

How similar are two plans?

Ordering of most-least similar plans and patients
Applications

• One framework
• Four questions:
  – Auto QA review rank priority
    • How complex is a plan?
  – Auto QA plan quality estimation
    • Does this plan make sense for this patient?
  – Auto Planning
    • What is the best plan for this patient?
Automated QA at Breast Cancer Rounds – A process to improve efficiency and quality of patient care

Dr. Kathy Rock
Breast Cancer Clinical Fellow
Rock K, Barry A, McIntosh C, Purdie T, Koch C.A
Distribution of Features

- Different plan complexities and qualities have different image and plan features
- Machine Learning learns hyperplanes to distinguish between the different complexity groups
- Learns to:
  - Automatically catch low quality plans
  - Rank plans in order of least-to-most complexity for review

Distribution of Features

- Complex plans
- Common Right Breast plan
- Common Left Breast Plan
- Poor quality plan
Plan Review Ranking

- Preliminary data: 375 breast RT plans (radical and boost)
- Plans manually scored for treatment complexity based on discussion requirements during rounds
  - Plans requiring more time given higher scores (higher review priority)
  - July 2014 to February 2015
- AutoQA learned to rank plans to match review discussion priority
- Performance measured with Normalized discounted cumulative gain
  - Score of 1.0 if the method perfectly ranks plans in agreement with review
  - AutoQA: 0.82
  - 4 field, 2 field, followed by boost: 0.77
  - Random: 0.61
Boost – complexity 0
Boost – complexity 2
Planning Error Detection

- Preliminary breast study, detects 80% of clinically rejected plans
- Detected plan error with poor high dose conformity (700 cGy isodose)
- Error not indicated by DVH alone

Rejected

Accepted
Automated Planning

- Typical pipeline:
  - Image $\rightarrow$ Contour $\rightarrow$ Beams $\rightarrow$ Optimization $\rightarrow$ Dose Map $\rightarrow$ QA
  - Many contours are entirely for plan optimization, not QA
- QA involves examining the dose distribution in the context of the image and ROIs to evaluate a plan
- Propose:
  - Use learned relationship of image characteristics and patient geometry to infer dose map per voxel
  - Image $\rightarrow$ Contour only QA structures $\rightarrow$ Dose Map $\rightarrow$ Beams $\rightarrow$ QA
DVH Drawbacks
The “Black Box”
Automated Treatment Plans
Opening the box

\[ P(dose \mid \text{Features}) \]

- Low density
- Gradient lrg. scale
- Inside lung
- Outside lung
- Gradient med. scale
- Inside heart
- Breast

\[ \text{Probability} \quad \text{Dose} \]

\[ \text{Probability} \quad \text{Dose} \]

\[ \text{Probability} \quad \text{Dose} \]

C. McIntosh and T. G. Purdie IEEE TMI 2015

chris.mcintosh@rmp.uhn.ca
Contextual Information

- Dose cannot be placed one voxel at a time like the digit recognition
- The distribution needs to make sense as a whole
- Contextual information models the relationship between similar patients
Automated Treatment Plans
Context From Experts

CT Image

Segmentation

Encoding (Density Estimation)
# Preliminary Statistics

<table>
<thead>
<tr>
<th>Site Group</th>
<th>Treatment Site</th>
<th>RT Technique</th>
<th>Dose Prescription (Total Dose in cGy, Number of Treatments)</th>
<th>Number of RT Plans For Dose Prediction</th>
<th>Predicted and Clinical Dose Distribution Concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td># Breast</td>
<td>Whole Breast</td>
<td>Tangent IMRT</td>
<td>4240_16</td>
<td>287</td>
<td>78.7</td>
</tr>
<tr>
<td># Breast</td>
<td>Cavity Boost</td>
<td>Non-Coplanar IMRT</td>
<td>1000_5</td>
<td>141</td>
<td>64.8</td>
</tr>
<tr>
<td># GU</td>
<td>Prostate</td>
<td>VMAT</td>
<td>7800_39</td>
<td>231</td>
<td>86.8</td>
</tr>
<tr>
<td>Head and Neck</td>
<td>Oropharynx</td>
<td>IMRT</td>
<td>7800_39</td>
<td>179</td>
<td>87.9</td>
</tr>
<tr>
<td>GI</td>
<td>Rectum</td>
<td>VMAT</td>
<td>5000_25</td>
<td>53</td>
<td>78.1</td>
</tr>
<tr>
<td>Lung</td>
<td>NSCLC</td>
<td>Sterestatic VMAT</td>
<td>4800_4</td>
<td>95</td>
<td>83.9</td>
</tr>
<tr>
<td>CNS Brain</td>
<td>Brainstem</td>
<td>IMRT</td>
<td>6000_30</td>
<td>168</td>
<td>87.4</td>
</tr>
<tr>
<td>CNS Brain</td>
<td>Brainstem</td>
<td>Protons</td>
<td>5400_30</td>
<td>10</td>
<td>92.7</td>
</tr>
<tr>
<td>CNS Brain</td>
<td>Post Fossa</td>
<td>Protons</td>
<td>3240_18</td>
<td>5</td>
<td>88.2</td>
</tr>
</tbody>
</table>

Total Plans: 1169

Score Mean: 81.8
Score Standard Deviation: 7.1

Abbreviations:
- CNS: Central Nervous System
- GU: Genitourinary
- GI: Gastrointestinal
- NSCLC: Non-Small Cell Lung Cancer
- IMRT: Intensity Modulated Radiation Therapy
- VMAT: Volumetric Modulated Arc Therapy

Note: Gamma Threshold set at 5% and 5 mm

# Plan Classes included in IEEE TMI paper

# Preliminary validation performed using leave-one-out due to smaller number of RT plans.

C. McIntosh and T. G. Purdie IEEE TMI 2015
<table>
<thead>
<tr>
<th>Site</th>
<th>Auto Plans Accepted</th>
<th>Clinical Plans Accepted</th>
<th>Head to Head Auto Win-Lose-Tie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prostate</td>
<td>4/5 SBOWEL dose</td>
<td>5/5</td>
<td>3-1-1 (80 %)</td>
</tr>
<tr>
<td>Stereotactic Lung</td>
<td>5/5</td>
<td>4/5 PTV coverage</td>
<td>2-1-2 (80 %)</td>
</tr>
<tr>
<td>Oropharynx</td>
<td>2/3 CORD dose</td>
<td>3/3</td>
<td>1-1-1 (67 %)</td>
</tr>
<tr>
<td>Rectum</td>
<td>1</td>
<td>0 High dose conformity</td>
<td>1-0-0 (100 %)</td>
</tr>
</tbody>
</table>
Blinded Review Inv. Plan

Auto or Clinical

A

B
Blinded Review Inv. Plan

Auto vs Clinical
Blinded Review Inv. Plan
Automated Spatial Dose Prediction Using Contextual Atlas Regression Forests

Presented by
Mattea Welch
PhD Candidate

Chris McIntosh, Mattea Welch, David A. Jaffray, and Thomas G. Purdie
Without OARs
Conclusions

• Going beyond the DVH
• Building patient and plan statistical relationships across a variety of data scales
• Answer questions relating:
  – Pairs of plans (planning complexity)
  – Pairs of patients and plans (plan quality)
  – Pairs of patients (Auto planning)
Acknowledgements

• This work would not be possible without the work and support of:
  – Dr. Tom Purdie
  – Dr. Tim Craig
  – Dr. Patricia Lindsay
  – Dr. Andrea McNiven
  – Igor Svistoun
  – Courtney Price
  – Andrea Marshall
Questions