Attention-gated convolutional neural networks for off-resonance correction of spiral real-time MRI

Yongwan Lim, Shrikanth S. Narayanan, Krishna S. Nayak

Ming Hsieh Department of Electrical and Computer Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, California, USA
Declaration of Financial Interests or Relationships

Speaker Name: Yongwan Lim

I have no financial interests or relationships to disclose with regard to the subject matter of this presentation.
Spiral Real-time MRI

Vocal tract

2.4mm², 12ms/frame, R=6.5
@ USC
Spiral Real-time MRI

Vocal tract

velum

2.4mm², 12ms/frame, R=6.5
@ USC

tlips
tongue
Spiral Real-time MRI

Vocal tract

- **Off-resonance artifacts** due to local susceptibility difference between air and tissue
  - Spatially and temporally varying

---

2.4mm², 12ms/frame, R=6.5

@ USC

---

Velum

Lips

Tongue
Spiral Real-time MRI

Vocal tract

- Off-resonance artifacts due to local susceptibility difference between air and tissue
  - Spatially and temporally varying

2.4mm², 12ms/frame, R=6.5 @ USC

lips

velum

tongue
**Vocal tract**

- **Off-resonance artifacts** due to local susceptibility difference between air and tissue
  - Spatially and temporally varying

**2.4mm², 12ms/frame, R=6.5 @ USC**

**Blurring Artifact**

**After De-Blurring**
Off-resonance Deblurring

• Standard Approaches\textsuperscript{1-4}:

Blurry Image

Deblurred Image

[1] KS Nayak et al, MRM. 2001
[2] BP Sutton et al, JMRI. 2010
[3] Y Lim et al. MRM. 2019
[4] DC Noll et al, MRM. 1992
Off-resonance Deblurring

• Standard Approaches\(^1\)\(^-\)\(^4\):

  1. Field map acquisition
     • Dual-TE (cf: single-TE or auto-focus)
     • Reduced scan efficiency

  2. Spatially-varying deconvolution
     • Non-iterative or iterative methods
     • Computationally slow (~minutes)

---

[1] KS Nayak et al, MRM. 2001
[2] BP Sutton et al, JMRI. 2010
[3] Y Lim et al. MRM. 2019
[4] DC Noll et al, MRM. 1992

---

Blurry Image  Field Map  Deblurred Image

---

YONGWANL@USC.EDU
Off-resonance Deblurring

- Standard Approaches\textsuperscript{1-4}:
  
  1. Field map acquisition
     - Dual-TE (cf: single-TE or auto-focus)
     - Reduced scan efficiency
  
  2. Spatially-varying deconvolution
     - Non-iterative or iterative methods
     - Computationally slow (~minutes)

[1] KS Nayak et al, MRM. 2001
[2] BP Sutton et al, JMRI. 2010
[3] Y Lim et al. MRM. 2019
[4] DC Noll et al, MRM. 1992
CNN-based Deblurring\textsuperscript{1}

[1] Y Lim, et al, MRM. 2020. 10.1002/mrm.28393
CNN-based Deblurring\textsuperscript{1}

Blurry image $\rightarrow$ 3-layer residual CNN $\rightarrow$ Deblurred Image

[1] Y Lim, et al, MRM. 2020. 10.1002/mrm.28393
CNN-based Deblurring\textsuperscript{1}

A supervised spatially varying deconvolution

In test time
1. Does NOT rely on field map
2. FAST (~milliseconds)

[1] Y Lim, et al, MRM. 2020. 10.1002/mrm.28393
Motivation

ReLU nonlinearity
• Provides a spatially-varying binary mask to convolution filters, enabling spatially-varying convolution.

[1] LC Man et al, MRM 1997 [2] DC Noll et al, MRM 1992
Motivation

ReLU nonlinearity

- Provides a spatially-varying binary mask to convolution filters, enabling spatially-varying convolution.

\[ M(F) = \begin{cases} 
1 & \text{if } F > 0 \\
0 & \text{o.w}
\end{cases} \]

[1] LC Man et al, MRM 1997 [2] DC Noll et al, MRM 1992
Motivation

ReLU nonlinearity

- Provides a spatially-varying binary mask to convolution filters, enabling spatially-varying convolution.

\[ M(F) = \begin{cases} 
1 & \text{if } F > 0 \\
0 & \text{o.w}
\end{cases} \]

\[ F' = F \otimes M(F) \]

[1] LC Man et al, MRM 1997 [2] DC Noll et al, MRM 1992
**Motivation**

ReLU nonlinearity

- Provides a spatially-varying binary mask to convolution filters, enabling spatially-varying convolution.

\[
M(F) = \begin{cases} 1 & \text{if } F > 0 \\ 0 & \text{o.w} \end{cases}
\]

\[F' = F \otimes M(F)\]

- The binary mask is computed only based on the **sign** of pixel value in an element-wise manner.

- It cannot exploit **local spatial or channel (filter) dependency**, unlike the conventional deblurring methods such as multi-frequency reconstruction\(^1\) or autofocus\(^2\).

---

\(^{1}\) LC Man et al, MRM 1997 \(^{2}\) DC Noll et al, MRM 1992
Goal of This Work

To exploit spatial and channel relationships of filtered outputs to improve the expressiveness of a network

…and enables an efficient off-resonance deblurring in the application of spiral RT-MRI of speech
Attention-gate CNN (AG-CNN)

Blury Input

9x9

F₁

5x5

F₁'

1x1

F₂'

Prediction

Depthwise separable conv 3X3 + ReLU
Depthwise separable conv 3X3 + sigmoid
Conv + tanh
Identity
Element-wise mul.
Element-wise add.
Attention-gate CNN (AG-CNN)

Attention Gate (AG) Module

Depthwise separable conv 3X3 + ReLU
Depthwise separable conv 3X3 + sigmoid
Conv + tanh
Identity
Element-wise mul.
Element-wise add.

Blurry Input

9x9

F_1

X

5x5

F_1'

F_2

M_1

M_2

F_2'

1x1

Prediction

\hat{Y}
Attention-gate CNN (AG-CNN)

Attention Gate (AG) Module

Depthwise separable conv 3X3 + ReLU
Depthwise separable conv 3X3 + sigmoid
Conv + tanh
Identity
Element-wise mul.
Element-wise add.
Attention-gate CNN (AG-CNN)

Attention Gate (AG) Module

Blurry Input

Depthwise separable conv 3X3 + ReLU
Depthwise separable conv 3X3 + sigmoid
Conv + tanh
Identity
Element-wise mul.
Element-wise add.

Predictions
Attention-gate CNN (AG-CNN)

Attention Gate (AG) Module

\[ F_1' = F_1 \otimes M_1(F_1) \]
\[ F_2' = F_2 \otimes M_2(F_2) \]

- Depthwise separable conv 3X3 + ReLU
- Depthwise separable conv 3X3 + sigmoid
- Conv + tanh
- Identity
- Element-wise mul.
- Element-wise add.

Blurry Input

\[ X \]

Prediction

\[ \hat{Y} \]
Attention-gate CNN (AG-CNN)

\[ F'_1 = F_1 \otimes M_1(F_1) \]
\[ F'_2 = F_2 \otimes M_2(F_2) \]
Methods

• **Data:**
  - 2D midsagittal speech spiral RT-MRI scans\(^1\)
  - Training data generation
    - Off-resonance correction\(^2\) and simulation\(^3\)

---

[1] S Lingala et al, MRM 2017
[2] Y Lim, et al, MRM 2019
[3] Y Lim et al, MRM 2020
[4] M Mathieu M, et al, ICLR 2015
[5] BP Sutton, et al, IEEE TMI 2003
Methods

**Data:**
- 2D midsagittal speech spiral RT-MRI scans
- Training data generation
- Train, validation, and test: 23, 5, and 5 subjects

**Network:**
- Loss function: $\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_{gdl}$ ($\mathcal{L}_{gdl}$: gradient difference loss)
- Adam optimizer, batch size = 64, epoch = 200

**Evaluation:**
- Comparisons: AG-CNN, CNN$^3$, IR (iterative reconstruction)
- Quality measures: PSNR, SSIM, HFEN

References:
[1] S Lingala et al, MRM 2017
[2] Y Lim, et al, MRM 2019
[3] Y Lim et al, MRM 2020
[4] M Mathieu M, et al, ICLR 2015
[5] BP Sutton, et al, IEEE TMI 2003
Results: Intermediate Layers

Blurry Input $X$

Attention Map

$F_1$

$M_1$

$F_1'$

Attention Map

$F_2$

$M_2$

$F_2'$

Prediction $Y$

Ground Truth $Y$

$\otimes$

$\oplus$

$9 \times 9$

$5 \times 5$

$1 \times 1$

$F_1$

$F_2$

$F_1'$

$F_2'$

$F_1$

$F_2$

$F_1'$

$F_2'$
Results: Intermediate Layers

Attention Map

Blurry Input $X$

$F_1$

$F_1'$

Attention Map

$M_1$

$F_2$

$M_2$

$F_2'$

Prediction $\hat{Y}$

Ground Truth $Y$
Results:
Intermediate Layers

Blurry Input $X$

Attention Map $M_1$

Attention Map $M_2$

Prediction $Y$

Ground Truth $\bar{Y}$
Results: Intermediate Layers

Blurry Input $X$ → $F_1$ → $M_1$ → $F'_1$ → $F_2$ → $M_2$ → $F'_2$ → Prediction $Y$

Attention Maps:
- $M_1$: Visualization of features $F_1$.
- $M_2$: Visualization of features $F_2$.

Ground Truth $Y$
Results:
Intermediate Layers

Blurry Input $X$

Attention Map $M_1$

F1

F1'

Attention Map $M_2$

F2

F2'

Prediction $\hat{Y}$

Ground Truth $Y$
### Results: Performance vs. Filter size

- Improved deblurring performance with less sensitivity to the kernel size but with a slight overhead.
- $(f_1, f_2) = (3, 3)$ is chosen.

| Architecture       | $(f_1, f_2)$ | Params | PSNR   | SSIM  | HFEN (x100) |
|---------------------|--------------|--------|--------|-------|-------------|
| CNN (9-5-1)         | -            | 61.7K  | 29.29  | 0.944 | 0.088       |
| +AG                 | (5,5)        | 70.7K  | 30.63  | 0.959 | 0.053       |
| +AG                 | (5,3)        | 70.0K  | 30.62  | 0.959 | 0.057       |
| +AG                 | (5,1)        | 69.6K  | 30.61  | 0.959 | 0.057       |
| +AG                 | (3,3)        | 68.4K  | **30.69** | 0.958 | 0.055       |
| +AG                 | (3,1)        | 68.1K  | 30.58  | 0.958 | 0.058       |
| (Blurred) Input     | -            | -      | 22.16  | 0.812 | 0.568       |
### Results: Performance vs. Filter size

- Improved deblurring performance with less sensitivity to the kernel size but with a slight overhead.
- \((f_1, f_2) = (3, 3)\) is chosen.

| Architecture     | \((f_1, f_2)\) | Params | PSNR  | SSIM  | HFEN (x100) |
|-------------------|----------------|--------|-------|-------|-------------|
| CNN (9-5-1)      | -              | 61.7K  | 29.29 | 0.944 | 0.088       |
| +AG              | (5,5)          | 70.7K  | 30.63 | 0.959 | 0.053       |
| +AG              | (5,3)          | 70.0K  | 30.62 | 0.959 | 0.057       |
| +AG              | (5,1)          | 69.6K  | 30.61 | 0.959 | 0.057       |
| +AG              | (3,3)          | 68.4K  | 30.69 | 0.958 | 0.055       |
| +AG              | (3,1)          | 68.1K  | 30.58 | 0.958 | 0.058       |
| (Blurred) Input  | -              | -      | 22.16 | 0.812 | 0.568       |
Results: Performance vs. Filter size

- Improved deblurring performance with less sensitivity to the kernel size but with a slight overhead.
- \((f_1, f_2) = (3, 3)\) is chosen.

| Architecture          | \((f_1, f_2)\) | Params | PSNR  | SSIM  | HFEN (x100) |
|-----------------------|----------------|--------|-------|-------|--------------|
| CNN (9-5-1)           | -              | 61.7K  | 29.29 | 0.944 | 0.088        |
| +AG                   | (5,5)          | 70.7K  | 30.63 | 0.959 | 0.053        |
| +AG                   | (5,3)          | 70.0K  | 30.62 | 0.959 | 0.057        |
| +AG                   | (5,1)          | 69.6K  | 30.61 | 0.959 | 0.057        |
| +AG                   | (3,3)          | 68.4K  | 30.69 | 0.958 | 0.055        |
| +AG                   | (3,1)          | 68.1K  | 30.58 | 0.958 | 0.058        |
| (Blurred) Input       | -              | -      | 22.16 | 0.812 | 0.568        |
Results: Performance vs. Filter size

- Improved deblurring performance with less sensitivity to the kernel size
  but with a slight overhead.
- \((f_1, f_2) = (3, 3)\) is chosen.

| Architecture      | \((f_1, f_2)\) | Params | PSNR  | SSIM | HFEN \(\times 100\) |
|-------------------|----------------|--------|-------|------|---------------------|
| CNN (9-5-1)       | -              | 61.7K  | 29.29 | 0.944| 0.088               |
| +AG               | (5,5)          | 70.7K  | 30.63 | 0.959| 0.053               |
| +AG               | (5,3)          | 70.0K  | 30.62 | 0.959| 0.057               |
| +AG               | (5,1)          | 69.6K  | 30.61 | 0.959| 0.057               |
| +AG               | (3,3)          | 68.4K  | **30.69** | 0.958| 0.055               |
| +AG               | (3,1)          | 68.1K  | 30.58 | 0.958| 0.058               |
| (Blurred) Input   | -              | -      | 22.16 | 0.812| 0.568               |

**filter size in AG module**
### Results: Performance vs. Filter size

- Improved deblurring performance with less sensitivity to the kernel size but with a slight overhead.
- \((f_1, f_2) = (3, 3)\) is chosen.

| Architecture          | \((f_1, f_2)\) | Params | PSNR  | SSIM  | HFEN \((x100)\) |
|-----------------------|----------------|--------|-------|-------|-----------------|
| CNN (9-5-1)           | -              | 61.7K  | 29.29 | 0.944 | 0.088           |
| +AG                   | (5, 5)         | 70.7K  | 30.63 | 0.959 | 0.053           |
| +AG                   | (5, 3)         | 70.0K  | 30.62 | 0.959 | 0.057           |
| +AG                   | (5, 1)         | 69.6K  | 30.61 | 0.959 | 0.057           |
| +AG                   | (3, 3)         | 68.4K  | 30.69 | 0.958 | 0.055           |
| +AG                   | (3, 1)         | 68.1K  | 30.58 | 0.958 | 0.058           |
| (Blurred) Input       | -              | -      | 22.16 | 0.812 | 0.568           |
Results: Comparisons

Graphs showing comparisons of PSNR, SSIM, and HFEN across different T_{read} [ms] values at 1.5T. The graphs compare Uncorrected, MFI w/ reference, IR w/ reference, CNN, and AG-CNN methods.
Results: Comparisons

![Graphs showing comparisons of PSNR, SSIM, and HFEN across different T_read values and techniques.](image)
Results: Test Data

Ground truth  Uncorrected

Error (X5)
Results: Test Data

Ground truth | Uncorrected | CNN | AG-CNN | IR with truth field map
---|---|---|---|---
PSNR: 20.12 | 27.02 | 29.47 | 37.39
Error (X5)
Results: Test Data

- Ground truth
- Uncorrected
- CNN
- AG-CNN
- IR with truth field map

PSNR: 20.12
27.02
29.47
37.39

Error (X5)
Results: Test Data

Ground truth  Uncorrected  CNN  AG-CNN  IR with truth field map

PSNR: 20.12  27.02  29.47  37.39

Error (X5)
Results: Test Data

Ground truth    Uncorrected    CNN    AG-CNN    IR with truth field map

PSNR: 20.12    27.02    29.47    37.39

USC Viterbi School of Engineering

YONGWANL@USC.EDU
Conclusion

• We develop the AG-CNN-based deblurring method for spiral RT-MRI in speech production.

• AG module could capture spatial and channel relationships of filtered outputs and improves deblurring performance with a slight overhead.

• An extensive comparison with existing attention approaches applicable to this task remains as future work.
Attention-gated convolutional neural networks for off-resonance correction of spiral real-time MRI

Yongwan Lim, Shrikanth S. Narayanan, Krishna S. Nayak

Thank you for your attention!

If you have any questions, please contact me: YONGWANL@USC.EDU