بسم اللّه الرحمن الرحيم

بررسی هرس شبکه عصبی در مسئله پرسش و پاسخ تصویری

پروژه کارشناسی مهندسی کامپیوتر

غزاله محمودی

استاد راهنما: دکتر سید صالح اعتمادی

دانشگاه علم و صنعت ایران

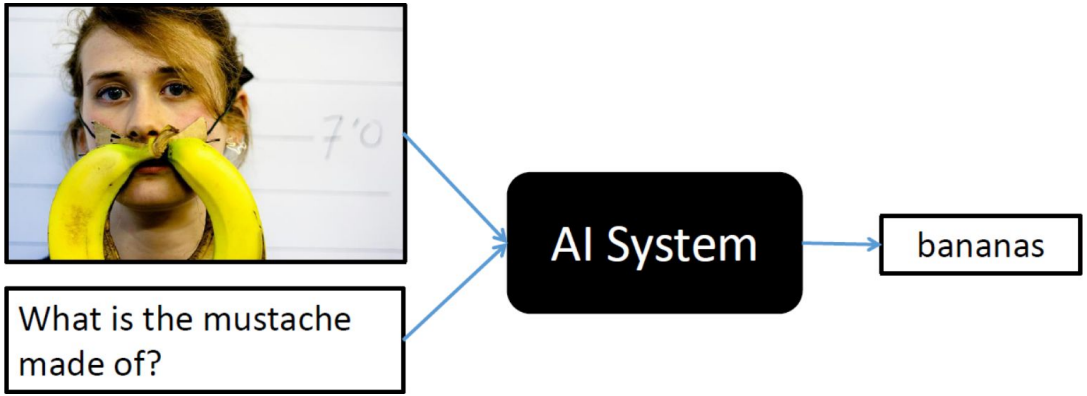
شهریور 1400

فهرست

1. مقدمه
2. مجموعه داده پرسش و پاسخ تصویری
3. شبکه LXMERT
4. فشرده‌سازی شبکه عصبی
5. نتیجه گیری
6. کارهای آینده

مقدمه

1. مسئله پرسش و پاسخ تصویری و کاربرد آن در زندگی.



مقدمه

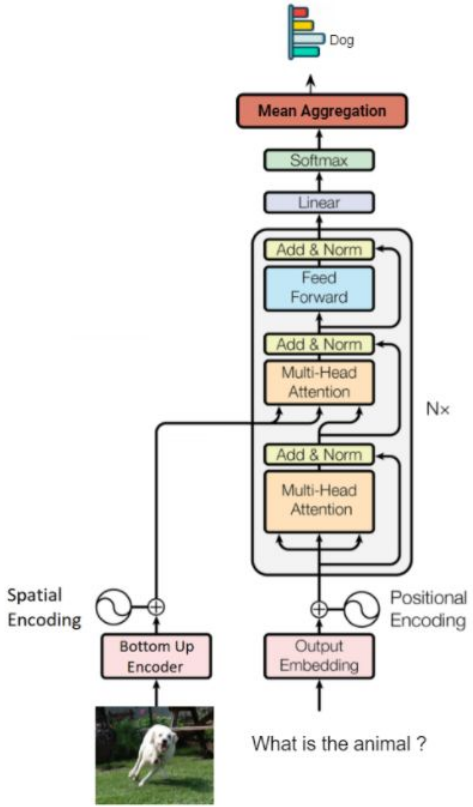
1. **مسئله پرسش و پاسخ تصویری و کاربرد آن در زندگی.**
   a. کمک به بهبود زندگی کم‌بینایان، نابینایان.

مقدمه

1. **مسئله پرسش و پاسخ تصویری و کاربرد آن در زندگی.**
   a. کمک به بهبود زندگی کم‌بینایان، نابینایان.
   b. کمک برای تفسیر تصاویر پیچیده پزشکی.

مقدمه

1. مسئله پرسش و پاسخ تصویری و کاربرد آن در زندگی.
2. **شبکه‌های ترنسفورمری برای حل مسئله.**



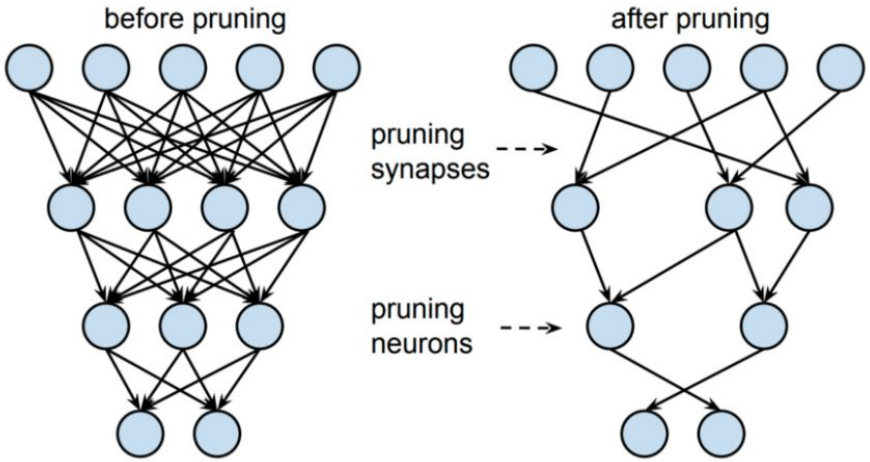
مقدمه

1. مسئله پرسش و پاسخ تصویری و کاربرد آن در زندگی.
2. شبکه‌های ترنسفورمری برای حل مسئله.
3. **چالش‌های شبکه‌های ترنسفورمری.**
    a. تعداد پارامترهای زیاد.

مقدمه

1. مسئله پرسش و پاسخ تصویری و کاربرد آن در زندگی.
2. شبکه‌های ترنسفورمری برای حل مسئله.
3. **چالش‌های شبکه‌های ترنسفورمری.**
    a. تعداد پارامترهای زیاد.
    b. محدودیت‌های پردازشی در تلفن همراه.

مقدمه



1. مسئله پرسش و پاسخ تصویری و کاربرد آن در زندگی.
2. شبکه‌های ترنسفورمری برای حل مسئله.
3. چالش‌های شبکه‌های ترنسفورمری.
4. **کاهش تعداد پارامترها با مکانیزم هرس شبکه عصبی.**

مجموعه داده پرسش و پاسخ تصویری

1. **مجموعه VQA v1.0.**

a. به طور میانگین به ازای هر تصویر 4 تا 5 پرسش.

b. به ازای هر پرسش 10 پاسخ از افراد منحصر به فرد.

	تعداد تصاویر	تعدادسوالات	تعداد پاسخ‌ها
داده‌های آموزشی	۸۲،۷۸۳	۲۴۸،۳۴۹	۲،۴۸۳،۴۹۰
داده‌های ارزیابی	۴۰،۵۰۴	۱۲۱،۵۱۲	۱،۲۱۵،۱۲۰
داده‌های تست	۸۱،۴۳۴	۲۴۴،۳۰۲	

مجموعه داده پرسش و پاسخ تصویری

1. **مجموعه VQA v1.0.**
   a. **تصاویر حقیقی.**
      i. برگرفته از مجموعه داده MSCOCO.

Q: What shape is the bench seat ?

A: oval, semi circle, curved, curved, double curve, banana, curved, wavy, twisting, curved

Q: What color is the stripe on the train ?

A: white, white, white, white, white, white, white, white, white, white

مجموعه داده پرسش و پاسخ تصویری

Q: Where are the flowers ?
A: near tree, tree, around tree, tree, by tree, around tree, around tree, grass, beneath tree, base of tree

Q: How many pillows ?
A: 1, 2, 2, 2, 2, 2, 2, 2, 2, 2

1. **مجموعه VQA v1.0.**

a. تصاویر حقیقی.

b. تصاویر انتزاعی.

i. داخل خانه.

ii. خارج خانه.

iii. 100 شی و 31 حیوان.

مجموعه داده پرسش و پاسخ تصویری

1. **مجموعه VQA v1.0.**
   a. تصاویر حقیقی.
   b. تصاویر انتزاعی.
   c. **نحوه جمع‌آوری مجموعه داده.**

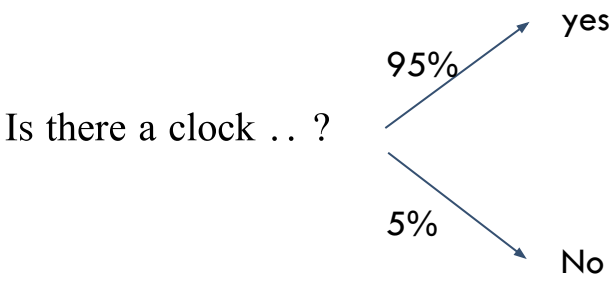
Q: Why are they standing?

(a) yes	**(b)** no		
(c) 1	**(d)** 2	**(e)** 3	**(f)** 4
(g) white	**(h)** red	**(i)** blue	**(j)** yellow
(k) playing game	**(l)** sheepskin	**(m)** waiting	**(n)** no where to sit
(o) firestone	**(p)** rugby	**(q)** forks	**(r)** waiting for train

Q: Is the TV on?

(a) yes	**(b)** no		
(c) 1	**(d)** 2	**(e)** 3	**(f)** 4
(g) white	**(h)** red	**(i)** blue	**(j)** yellow
(k) shag	**(l)** jeopardy	**(m)** sports	**(n)** between big elephants
(o) edinburgh	**(p)** strawberries	**(q)** tv show	**(r)** white streak on face

مجموعه داده پرسش و پاسخ تصویری



1. **مجموعه VQA v1.0.**
   a. چالش
      i. وجود تعصبات زبانی

مجموعه داده پرسش و پاسخ تصویری

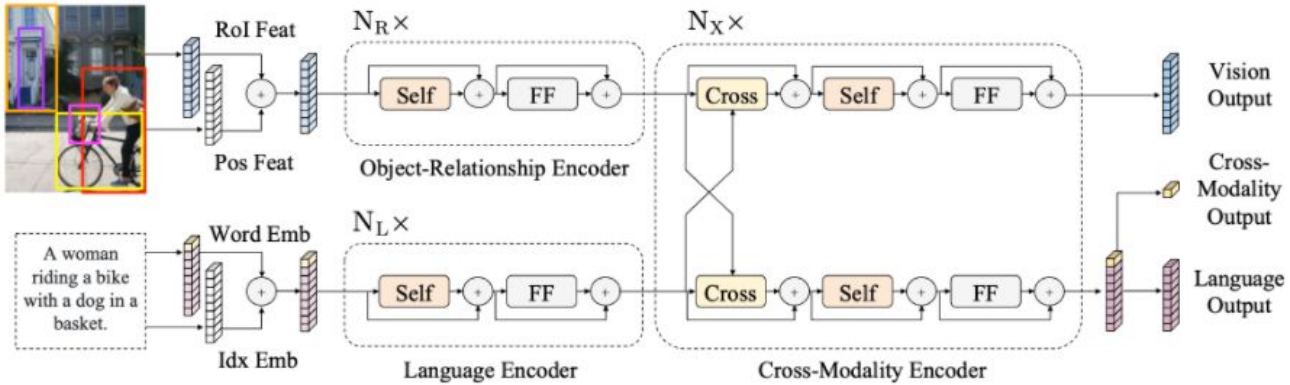
1. مجموعه VQA v1.0.
2. **مجموعه VQA v2.0.**
    a. بهبود مشکل تعصبات زبانی در مجموعه داده پیشین.

مجموعه داده پرسش و پاسخ تصویری

1. مجموعه VQA v1.0.
2. **مجموعه VQA v2.0.**
   a. بهبود مشکل تعصبات زبانی در مجموعه داده پیشین.
   b. جمع‌آوری تصاویر مکمل.

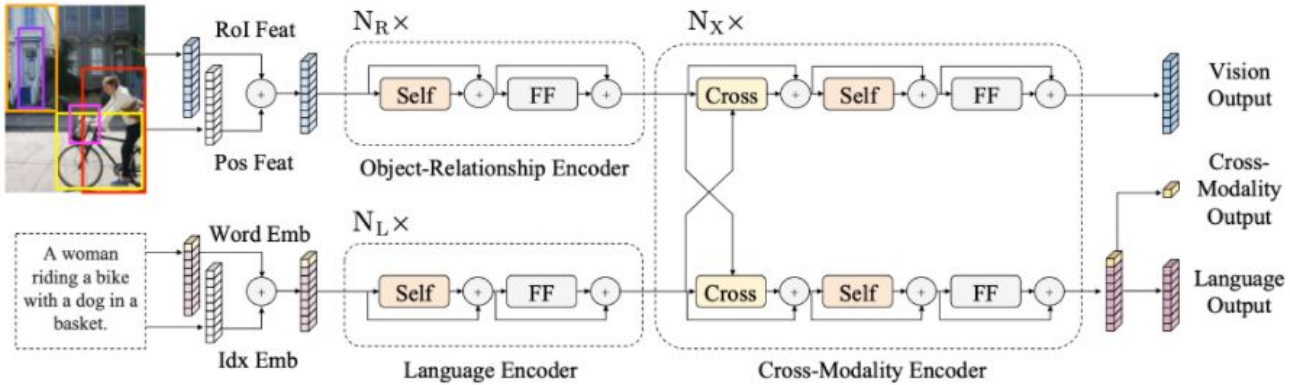
شبکه LXMERT

- **معماری شبکه.**
  - شبکه ترنسفورمری دوجریانه.



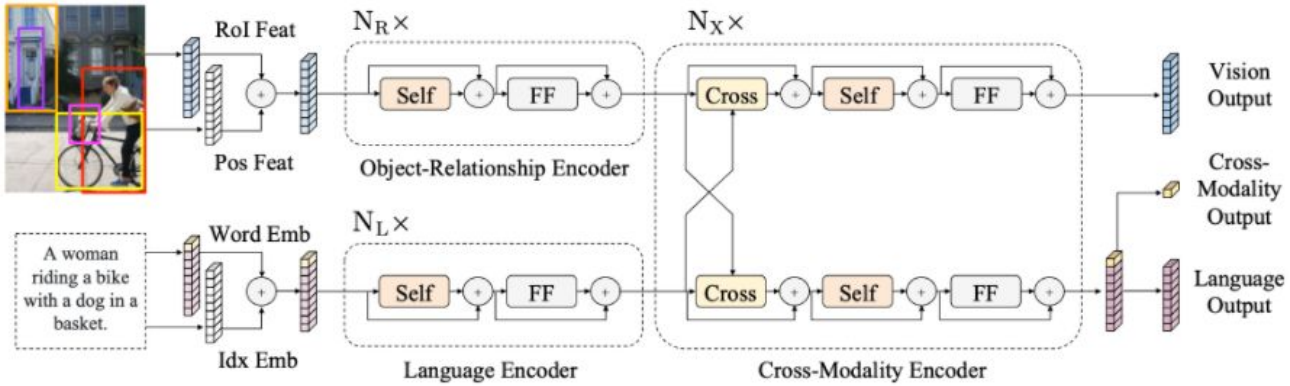
شبکه LXMERT

- **معماری شبکه.**
  - شبکه ترنسفورمری دوجریانه.
  - ورودی
    - کلمات جمله
    - RoI تصویر



شبکه LXMERT

- **معماری شبکه.**
  - شبکه ترنسفورمری دوجریانه.
  - ورودی
    - کلمات جمله
    - RoI تصویر
  - خروجی
    - بخش Vision
    - بخش Language
    - بخش Cross-Modality (مشابه توکن cls در شبکه BERT)



فشرده‌سازی شبکه LXMERT

1. **هرس بر اساس وزن اتصالات (Magnitude Pruning).**
    a. **هرس اتصالات با وزن کم.**

فشرده‌سازی شبکه LXMERT

1. **هرس بر اساس وزن اتصالات (Magnitude Pruning).**
    a. **هرس اتصالات با وزن کم.**
        i. **فرضیه بلیت بخت‌آزمایی.**

1. یک شبکه کاملا متصل که به صورت تصادفی مقدار دهی اولیه شده است شامل یک زیر شبکه می‌باشد به طوری که اگر آن زیر شبکه را به تعداد تکرار مشابه شبکه اصلی آموزش دهیم، دقت روی داده تست در هر دو حالت تقریبا یکسان خواهد شد.

فشرده‌سازی شبکه LXMERT

1. **هرس بر اساس وزن اتصالات (Magnitude Pruning).**
   a. هرس اتصالات با وزن کم.
   b. هرس اتصالات به صورت تصادفی.

فشرده‌سازی شبکه LXMERT

1. **هرس بر اساس وزن اتصالات (Magnitude Pruning).**
   a. هرس اتصالات با وزن کم.
   b. هرس اتصالات به صورت تصادفی.
   c. هرس اتصالات با وزن زیاد.

فشرده‌سازی شبکه LXMERT

1. هرس بر اساس وزن اتصالات (Magnitude Pruning).
    a. هرس اتصالات با وزن کم.
    b. هرس اتصالات به صورت تصادفی.
    c. هرس اتصالات با وزن زیاد.
2. **به ازای انواع هرس از 10 درصد تا 90 درصد وزن‌های شبکه هرس شدند.**

فشرده‌سازی شبکه LXMERT

1. هرس بر اساس وزن اتصالات (Magnitude Pruning).
    a. هرس اتصالات با وزن کم.
    b. هرس اتصالات به صورت تصادفی.
    c. هرس اتصالات با وزن زیاد.
2. به ازای انواع هرس از 10 درصد تا 90 درصد وزن‌های شبکه هرس شدند.
3. **برای اطمینان از نتایج آزمایش‌ها در سه seed تکرار شدند.**

هرس اتصالات با وزن کم

1. **این روش همان فرضیه بلیت بخت‌آزمایی می‌باشد.**

هرس اتصالات با وزن کم

1. این روش همان فرضیه بلیت بخت‌آزمایی می‌باشد.
2. **روش حل مسئله VQA از نوع رده‌بندی است.**
    a. 3129 جواب پرتکرار.

مراحل هرس اتصالات با وزن کم

1. وزن‌های از قبل آموزش دیده و رده‌بند VQA به شبکه LXMERT داده می‌شود.

مراحل هرس اتصالات با وزن کم

1. وزن‌های از قبل آموزش دیده و رده‌بند VQA به شبکه LXMERT داده می‌شود.
2. fine-tune کردن شبکه بر روی مجموعه داده VQA v2.0.
   a. Unpruned Baseline

مراحل هرس اتصالات با وزن کم

1. وزن‌های از قبل آموزش دیده و رده‌بند VQA به شبکه LXMERT داده می‌شود.
2. fine-tune کردن شبکه بر روی مجموعه داده VQA v2.0.
    a. Unpruned Baseline
3. به صورت تکرار شونده در هر مرحله 10 درصد اتصالات با وزن کم را حذف می‌شود.
    a. به جز اتصالات لایه embedding و لایه خروجی.
    b. Pruned

مراحل هرس اتصالات با وزن کم

1. وزن‌های از قبل آموزش دیده و رده‌بند VQA به شبکه LXMERT داده می‌شود.
2. fine-tune کردن شبکه بر روی مجموعه داده VQA v2.0.
    a. Unpruned Baseline
3. به صورت تکرار شونده در هر مرحله 10 درصد اتصالات با وزن کم را حذف می‌شود.
    a. به جز اتصالات لایه embedding و لایه خروجی.
    b. Pruned
4. پس از حذف اتصالات وزن‌های از قبل آموزش دیده اولیه به شبکه بازنشانی می‌شود.
    a. reset initial weight

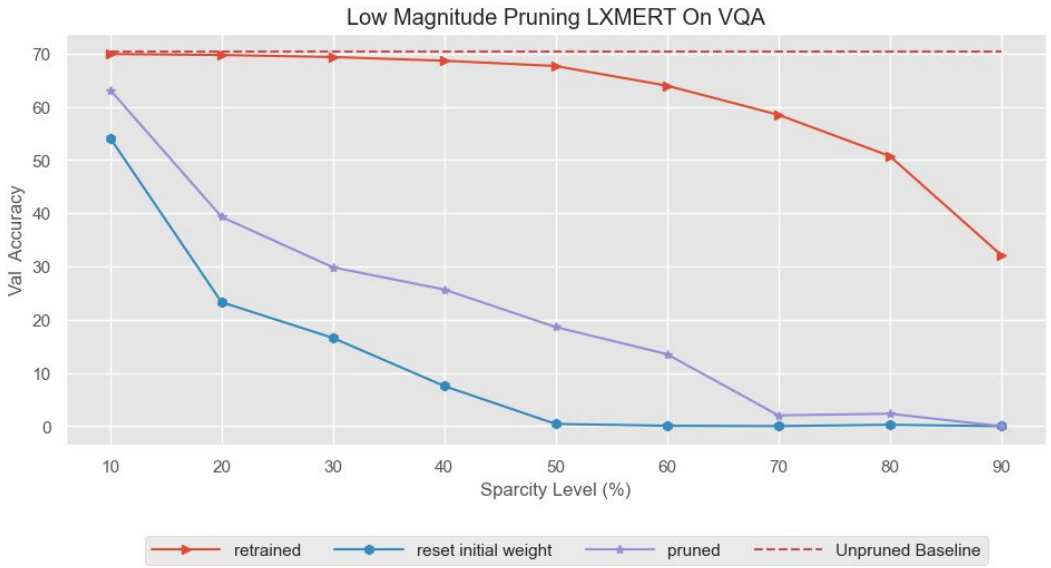
مراحل هرس اتصالات با وزن کم

1. وزن‌های از قبل آموزش دیده و رده‌بند VQA به شبکه LXMERT داده می‌شود.
2. fine-tune کردن شبکه بر روی مجموعه داده VQA v2.0.
    a. Unpruned Baseline
3. به صورت تکرار شونده در هر مرحله 10 درصد اتصالات با وزن کم را حذف می‌شود.
    a. به جز اتصالات لایه embedding و لایه خروجی.
    b. Pruned
4. پس از حذف اتصالات وزن‌های از قبل آموزش دیده اولیه به شبکه بازنشانی می‌شود.
    a. reset initial weight
5. آموزش مجدد شبکه هرس شده بر مجموعه داده VQA v2.0.
    a. retrain

نتایج هرس اتصالات

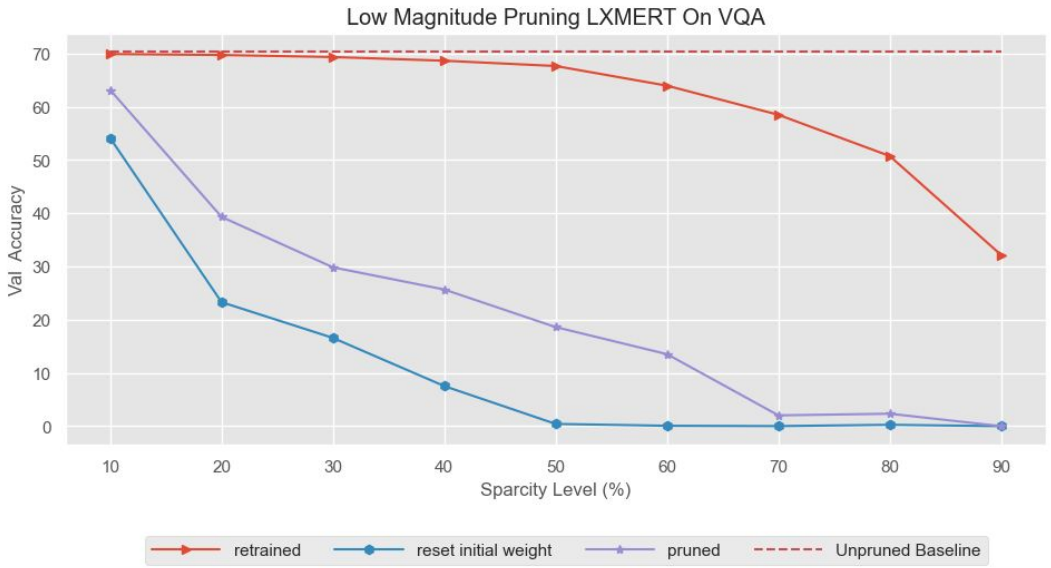
1. **هرس اتصالات با وزن کم.**

نتایج هرس اتصالات



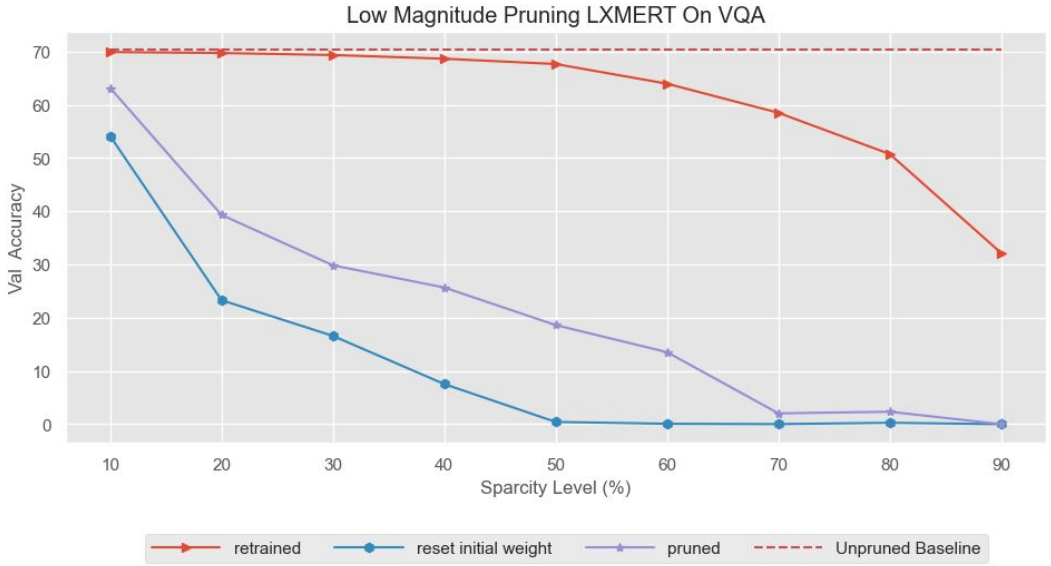
1. **هرس اتصالات با وزن کم.**

نتایج هرس اتصالات



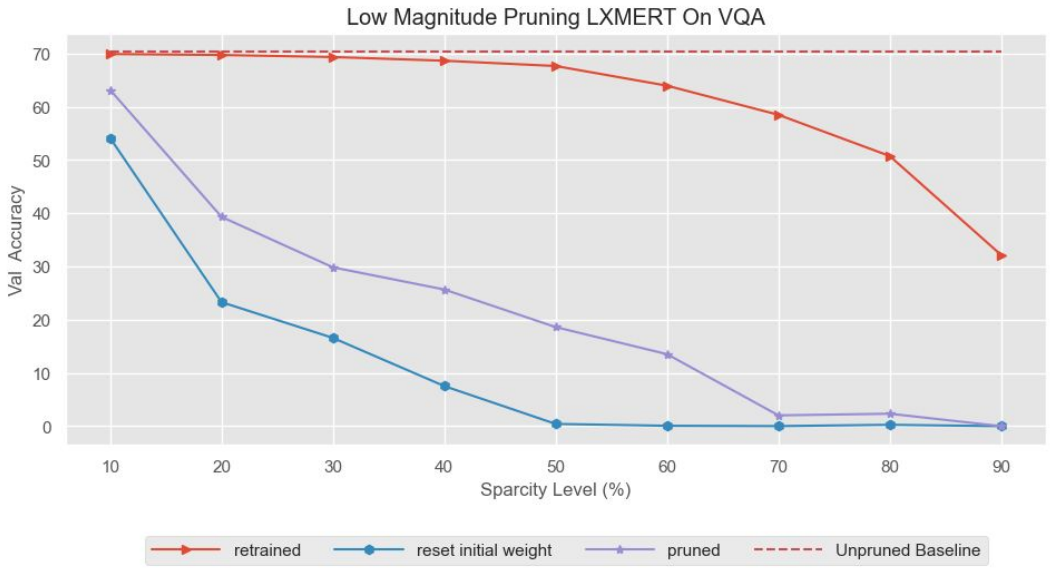
1. **هرس اتصالات با وزن کم.**
   a. **حذف 40 - 50 درصد اتصالات با تاثیر کم بر عملکرد نهایی شبکه.**

نتایج هرس اتصالات



1. **هرس اتصالات با وزن کم.**
   a. حذف 40 - 50 درصد اتصالات با تاثیر کم بر عملکرد نهایی شبکه.
   b. **صحت فرضیه بلیت بخت‌آزمایی در شبکه.**

نتایج هرس اتصالات

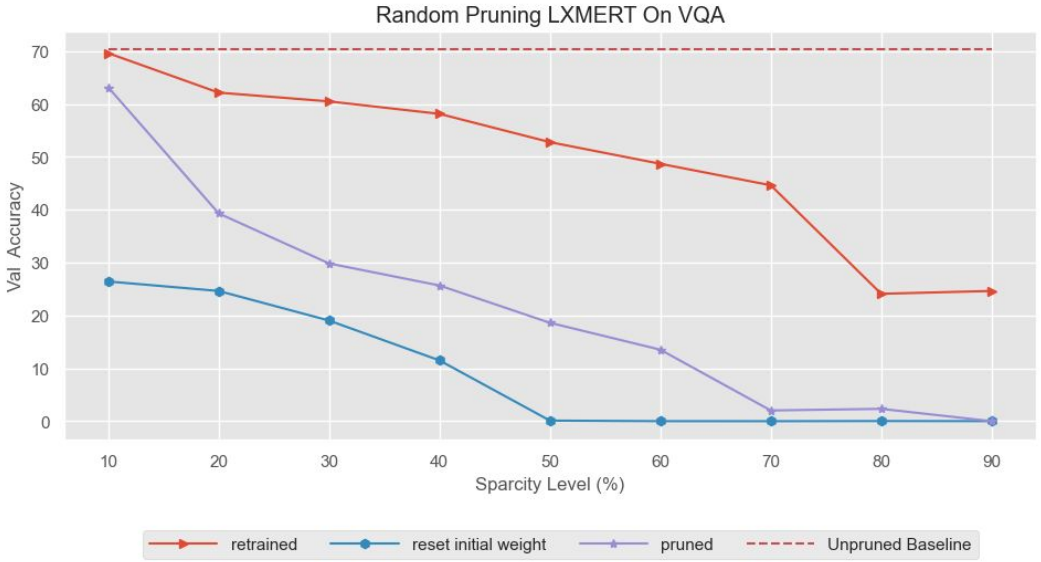


1. **هرس اتصالات با وزن کم.**
   a. حذف 40 - 50 درصد اتصالات با تاثیر کم بر عملکرد نهایی شبکه.
   b. صحت فرضیه بلیت بخت‌آزمایی در شبکه.
   c. **اتصالات کم‌وزن تاثیر کمتری در عملکرد شبکه دارند.**

نتایج هرس اتصالات

1. هرس اتصالات با وزن کم.
2. **هرس اتصالات به صورت تصادفی.**

نتایج هرس اتصالات

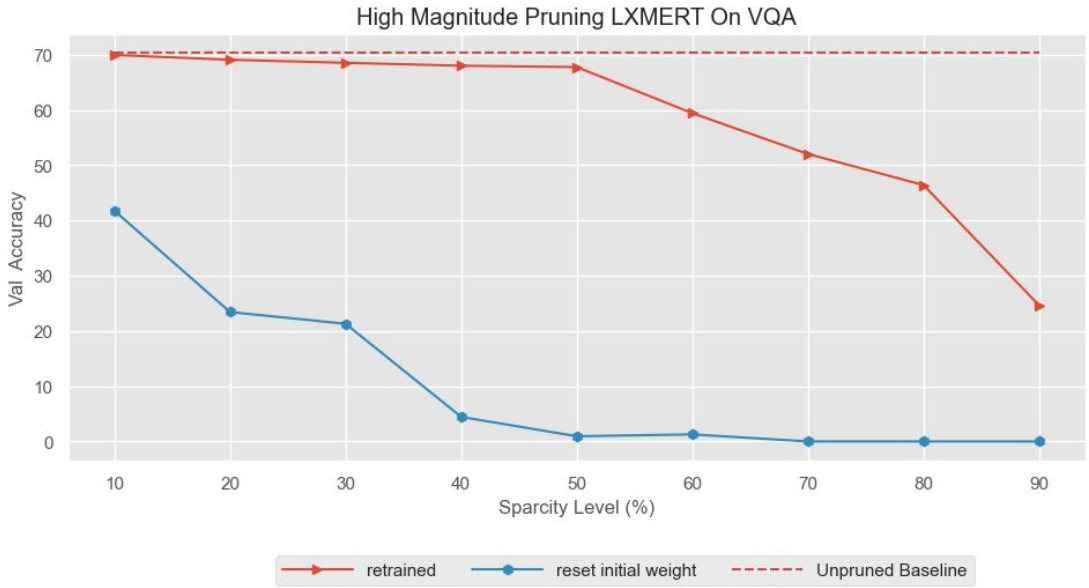


1. هرس اتصالات با وزن کم.
2. **هرس اتصالات به صورت تصادفی.**

نتایج هرس اتصالات

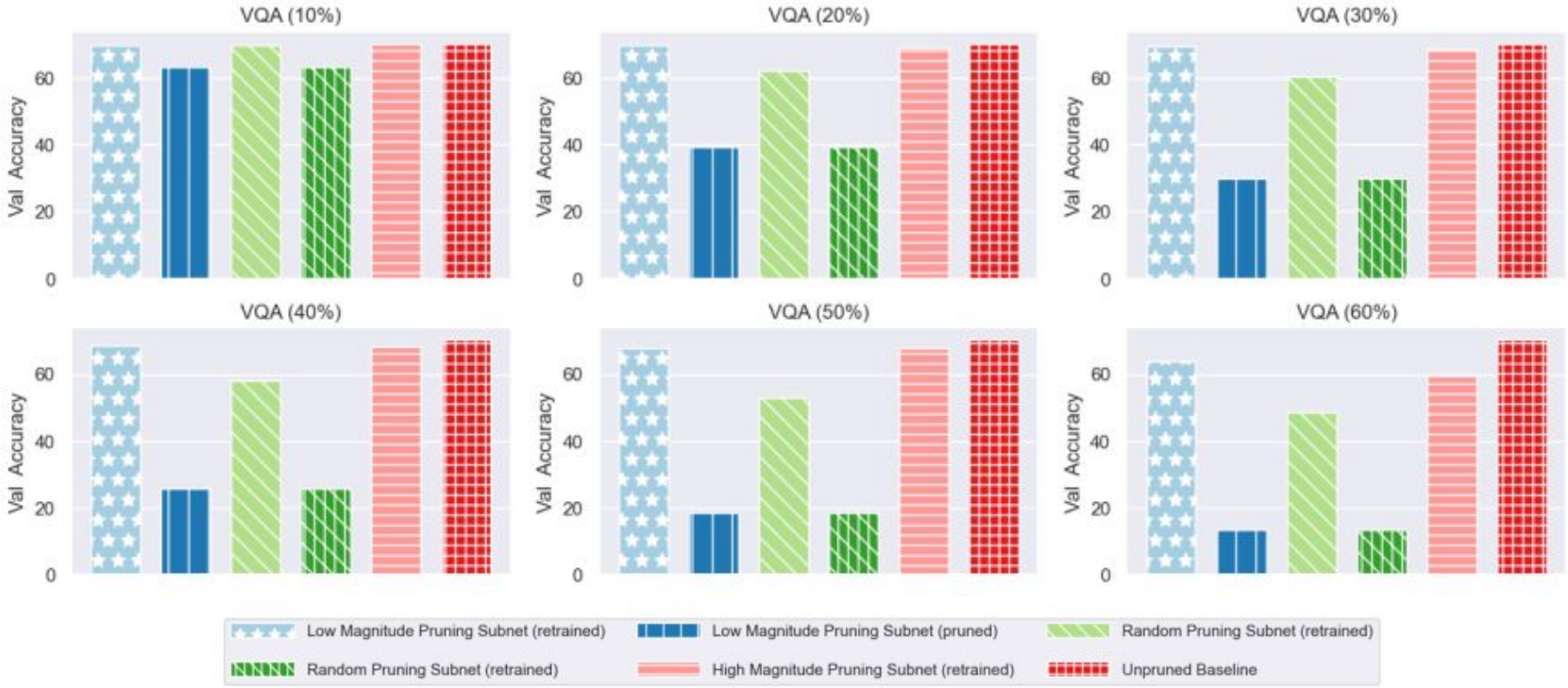
1. هرس اتصالات با وزن کم.
2. هرس اتصالات به صورت تصادفی.
3. **هرس اتصالات با وزن زیاد.**

نتایج هرس اتصالات



1. هرس اتصالات با وزن کم.
2. هرس اتصالات به صورت تصادفی.
3. **هرس اتصالات با وزن زیاد.**

نتایج هرس اتصالات



نتایج هرس اتصالات

درصد هرس	اتصالات با وزن کم	اتصالات تصادفی	اتصالات با وزن زیاد
۱۰	69.94 ± 0.03	69.58 ± 0.02	69.75 ± 0.13
۲۰	69.59 ± 0.11	62.21 ± 0.06	69.17 ± 0.09
۳۰	69.23 ± 0.07	60.23 ± 0.24	68.60 ± 0.08
۴۰	68.70 ± 0.05	57.49 ± 0.62	67.87 ± 0.18
۵۰	67.44 ± 0.14	52.63 ± 0.13	67.78 ± 0.04
۶۰	63.94 ± 0.01	48.68 ± 0.00	58.76 ± 0.63
۷۰	58.50 ± 0.04	44.96 ± 0.34	52.28 ± 0.26
۸۰	50.63 ± 0.21	24.36 ± 0.26	46.54 ± 0.15
۹۰	28.12 ± 4.02	24.36 ± 0.26	25.40 ± 0.77

نتیجه‌گیری

- امکان حذف 50 درصد از اتصالات کم‌وزن شبکه LXMERT با کمترین تاثیر در عملکرد شبکه در مسئله پرسش و پاسخ تصویری وجود دارد. دقت این حالت به 95% دقت مدل کامل دست می‌یابد.
- بهترین نوع هرس از بین موارد آزمایش شده، هرس اتصالات با وزن کم می‌باشد.
- فرضیه بلیت بخت‌آزمایی برای شبکه دوجریانه LXMERT بر مسئله VQA برقرار است.

کارهای آینده

- بررسی نتایج هرس عصبی شبکه LXMERT بر مجموعه داده GQA و NLVR2.
- بررسی روش‌های دیگر هرس و تاثیر آن بر زمان آموزش شبکه.
- بررسی دقیق‌تر معماری LXMERT برای توجیه بهتر نتایج حاصل از هرس اتصالات با وزن زیاد.
- بررسی عملی شبکه فشرده شده در برنامه‌های کاربردی موبایل.
- گردآوری مجموعه داده فارسی برای مسئله پرسش و پاسخ تصویری.

با تشکر از توجه شما

پیاده‌سازی پروژه: https://github.com/ghazaleh-mahmoodi/lxmert-compression

gh_mhdi@outlook.com

منابع

- ANTOL, S., AGRAWAL, A., LU, J., MITCHELL, M., BATRA, D., LAWRENCE ZITNICK, C., AND PARIKH, D. Vqa: Visual question answering. in Proceedings of the IEEE international conference on computer vision (2015), pp. 2425–2433
- TAN, H. H., AND BANSAL, M. Lxmert: Learning cross-modality encoder representations from transformers. in EMNLP/IJCNLP (2019)
- FRANKLE, J., AND CARBIN, M. The lottery ticket hypothesis: Finding sparse, trainable neural networks, 2019.
- GAN, Z., CHEN, Y.-C., LI, L., CHEN, T., CHENG, Y., WANG, S., AND LIU, J. Playing lottery tickets with vision and language, 2021.
- CHEN, Y.-C., LI, L., YU, L., KHOLY, A. E., AHMED, F., GAN, Z., CHENG, Y., AND LIU, J. Uniter: Universal image-text representation learning, 2020.
- LIN, T.-Y., MAIRE, M., BELONGIE, S., BOURDEV, L., GIRSHICK, R., HAYS, J., PERONA, P.,RAMANAN, D., ZITNICK, C. L., AND DOLLÁR, P. Microsoft coco: Common objects in context, 2015.
- DEVLIN, J., CHANG, M.-W., LEE, K., AND TOUTANOVA, K. BERT: Pre-training of deep bidirectional transformers for language understanding. in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (Minneapolis, Minnesota, June 2019), Association for Computational Linguistics, pp. 4171–4186.
- BEN ABACHA, A., HASAN, S. A., DATLA, V. V., LIU, J., DEMNER-FUSHMAN, D., AND MÜLLER, H. VQA-Med: Overview of the medical visual question answering task at imageclef 2019. In CLEF2019 Working Notes (Lugano, Switzerland, September 09-12 2019), CEUR Workshop Proceedings, CEUR-WS.org ‹http://ceur-ws.org›.
- GURARI, D., LI, Q., STANGL, A. J., GUO, A., LIN, C., GRAUMAN, K., LUO, J., AND BIGHAM, J. P. Vizwiz grand challenge: Answering visual questions from blind people, 2018.