

پیشینه تحقیق

استاد راهنما  
سرکار خانم دکتر مریم عابدی

دانشجو  
محمد شاهپوری

آذر  
۱۴۰۱

## فهرست مطالب

ii

فهرست تصاویر

iii

فهرست جداول

۱	مقدمه	۱
۱	تشخیص شی برجسته چیست؟	۱.۱
۱	تاریخچه تشخیص شی برجسته	۲.۱
۲	مدل‌های تشخیص شی برجسته	۲
۲	مدل‌های قدیمی	۱.۲
۲	مدل‌های مبتنی بر بلوک با نشانه‌های درونی	۱.۱.۲
۳	مدل‌های مبتنی بر منطقه با نشانه‌های درونی	۲.۱.۲
۳	مدل‌ها با نشانه‌های بیرونی	۳.۱.۲
۳	مدل‌های جدید	۲.۲
۳	شبکه‌های عصبی کانولوشنال	۱.۲.۲
۳	شبکه‌های عصبی کاملاً کانولوشنال	۲.۲.۲
۳	کاربردهای تشخیص شی برجسته	۳
۴	مجموعه داده‌ها و معیارهای ارزیابی	۴
۴	معیارهای ارزیابی	۱.۴
۴	مجموعه داده‌ها	۲.۴
۵	مراجع	

## فهرست تصاویر

۱	تعدادی از کاربردهای تشخیص شی برجسته	۴
---	-------------------------------------	---

## فهرست جداول

## ۱ مقدمه

انسان‌ها قادرند مناطق متمایز درون تصویر را به سرعت و به راحتی تشخیص دهند. ولی در سیستم‌های کامپیوتری برای بدست آوردن اطلاعات سطح بالا باید بر روی تصویر محاسبات و پردازش‌هایی انجام شود. برخی از موضوعاتی که به صورت مستقیم یا غیر مستقیم با برجستگی بصری<sup>۱</sup> در ارتباط هستند عبارتند از: پیش بینی تثبیت<sup>۲</sup> [۱، ۲]، اهمیت شی<sup>۳</sup> [۳، ۴، ۵]، به یاد ماندنی<sup>۴</sup> [۶]، بهم ریختگی صحنه<sup>۵</sup> [۷]، جذابیت ویدیویی<sup>۶</sup> [۸، ۹، ۱۰، ۱۱]، تعجب<sup>۷</sup> [۱۲]، ارزیابی کیفیت تصویر<sup>۸</sup> [۱۳، ۱۴، ۱۵]، معمولی بودن صحنه<sup>۹</sup> [۱۶، ۱۷]، زیبایی شناختی<sup>۱۰</sup> [۸] و ویژگی‌ها<sup>۱۱</sup> [۱۸].

### ۱.۱ تشخیص شی برجسته چیست؟

«تشخیص شی برجسته»<sup>۱۲</sup> یا «بخش بندی شی برجسته»<sup>۱۳</sup> در بینایی کامپیوتر به دو بخش تقسیم می‌شود: (۱) تشخیص برجسته‌ترین شی (۲) بخش بندی دقیق منطقه آن شی. با این حال بیشتر مدل‌ها بین این دو مرحله تمایز چندانی قائل نمی‌شوند. با دنبال کردن تحقیقات اولیه که توسط Itti و همکاران [۱۹] و Liu و همکاران [۲۰] انجام شده است در میابیم که مدل‌ها هر دو مرحله را همزمان انجام می‌دهند.

### ۲.۱ تاریخچه تشخیص شی برجسته

تشخیص شی برجسته را می‌توان به سه مقطع زمانی تقسیم بندی کرد. اولین آن‌ها که با کار Itti و همکاران [۱۹] در سال ۱۹۹۸ شروع می‌شود. این مدل از اولین روش‌های محاسباتی پایین به بالا بر اساس مرکز محصور شده<sup>۱۴</sup> است. در [۱۹] Itti و همکاران مثال‌هایی را نمایش داده‌اند که مدل آن‌ها قادر به تشخیص ناپیوستگی‌های فضایی<sup>۱۵</sup> در تصویر است.

دومین مقطع زمانی از سال ۲۰۰۷ با منطبق کردن مسئله تشخیص شی برجسته به مسئله تقسیم بندی دودویی<sup>۱۶</sup> شروع شد. Liu و همکاران [۲۰، ۲۱] و Achanta و همکاران [۲۲] اولین کارها در این مقطع زمانی بوده‌اند. این محققان از برخی مدل‌های قبلی که در تلاش برای تشخیص شی برجسته بوده‌اند الهام گرفته‌اند مانند: Ma و Zhang [۲۳]، Liu و Gleicher [۲۴] و Walther و همکاران [۲۵].

- 
- Visual saliency ۱
  - Fixation prediction ۲
  - Object importance ۳
  - Memorability ۴
  - Scene clutter ۵
  - Video interestingness ۶
  - Surprise ۷
  - Image quality assessment ۸
  - Scene typicality ۹
  - Aesthetic ۱۰
  - Attributes ۱۱
  - Salient object detection ۱۲
  - Salient object segmentation ۱۳
  - Center-surround ۱۴
  - Spatial discontinuities ۱۵
  - Binary segmentation problem ۱۶

مقطع زمانی سوم با استفاده از شبکه‌های عصبی کانولوشنال<sup>۱</sup> [۲۶]، به صورت خاص شبکه‌های عصبی کاملاً کانولوشنال<sup>۲</sup> [۲۷]، در تشخیص شی برجسته از سال ۲۰۱۵ شروع شد. برخلاف بسیاری از روش‌های گذشته که بر اساس محاسبه کنتراست بوده‌اند، روش‌های مبتنی بر شبکه‌های عصبی کانولوشنال نیاز به ویژگی‌های دست ساز<sup>۳</sup> را از بین برده‌اند، وابستگی به دانش سوگیری مرکز<sup>۴</sup> را کاهش داده‌اند و توسط بسیاری از محققین پذیرفته شده‌اند.

## ۲ مدل‌های تشخیص شی برجسته

مدل‌های ارائه شده را می‌توان به دو دسته قدیمی و جدید تقسیم کرد. مدل‌های قدیمی مدل‌های محاسباتی<sup>۵</sup> و مدل‌های جدید مدل‌های شبکه‌های عصبی هستند.

### ۱.۲ مدل‌های قدیمی

تعداد زیادی از مدل‌هایی که در دو دهه گذشته برای تشخیص شی برجسته ارائه شده‌اند. هدف بیشتر این روش‌ها شناسایی زیرمجموعه‌هایی<sup>۶</sup> از شی برجسته است و سپس با ادغام کردن آن‌ها، شی برجسته را از پس زمینه جدا می‌کنند.

به صورت کلی می‌توان مدل‌های قدیمی را بر اساس نوع عملیات یا ویژگی‌هایی که مورد استفاده قرار می‌دهند دسته بندی کرد.

(۱) تجزیه و تحلیل مبتنی بر بلوک و تجزیه و تحلیل مبتنی بر منطقه: برای تشخیص شی برجسته دو نوع از زیر مجموعه‌های بصری با نام‌های بلوک<sup>۷</sup> و منطقه<sup>۸</sup> مورد استفاده قرار می‌گیرند.

(۲) نشانه‌های درونی و نشانه‌های بیرونی: نشانه‌های درونی<sup>۹</sup> با استفاده از محاسباتی بر روی تصویر سعی در پیدا کردن شی برجسته می‌کنند. در حالی که نشانه‌های بیرونی<sup>۱۰</sup> مانند نشانه گذاری‌های کاربر<sup>۱۱</sup>، نقشه عمق<sup>۱۲</sup> یا اطلاعات آماری از تصاویر مشابه<sup>۱۳</sup> در کنار نشانه‌های درونی برای تشخیص شی برجسته مورد استفاده قرار می‌گیرند.

#### ۱.۱.۲ مدل‌های مبتنی بر بلوک با نشانه‌های درونی

در این مدل‌ها نشانه‌های درونی برای تشخیص شی برجسته از بلوک‌ها استخراج شده‌اند. [۱۹، ۲۲، ۲۴، ۲۸، ۲۹، ۳۰، ۳۱، ۳۲، ۳۳، ۳۴، ۳۵]

۱ Convolutional neural networks (CNNs)

۲ Fully convolutional neural networks

۳ Hand-crafted

۴ Center bias knowledge

۵ Computational

۶ زیر مجموعه‌های بصری می‌توانند پیکسل‌ها، بلوک‌ها، سوپرپیکسل‌ها و مناطق باشند.

۷ Block

۸ Region

۹ Intrinsic cues

۱۰ Extrinsic cues

۱۱ User annotations

۱۲ Depth map

۱۳ Statistical information of similar images

## ۲.۱.۲ مدل‌های مبتنی بر منطقه با نشانه‌های درونی

برخلاف دسته قبلی که نشانه‌های درونی از بلوک‌ها استخراج می‌شدند در این دسته نشانه‌ها از مناطق استخراج می‌شوند. [۳۷، ۳۶، ۲۴، ۳۸، ۳۹، ۴۰، ۴۱، ۳۵]

## ۳.۱.۲ مدل‌ها با نشانه‌های بیرونی

در دسته سوم مدل‌هایی قرار می‌گیرند که از نشانه‌های بیرونی نیز برای تشخیص شی برجسته استفاده کرده‌اند. [۴۴، ۴۳، ۴۲]

## ۲.۲ مدل‌های جدید

مدل‌های جدید مدل‌هایی هستند که از شبکه‌های عصبی کانولوشنال و شبکه‌های عصبی کاملاً کانولوشنال در آن‌ها برای تشخیص شی برجسته استفاده شده است.

### ۱.۲.۲ شبکه‌های عصبی کانولوشنال

در دسته اول از این بخش، از شبکه‌های عصبی کانولوشنال برای تشخیص شی برجسته استفاده شده است. [۴۸، ۴۷، ۴۶، ۴۵]

### ۲.۲.۲ شبکه‌های عصبی کاملاً کانولوشنال

شبکه‌های عصبی کاملاً کانولوشنال را می‌توان در دسته دوم قرار داد که سعی در تشخیص شی برجسته می‌کنند. [۵۲، ۵۱، ۵۰، ۴۹]

## ۳ کاربردهای تشخیص شی برجسته

تشخیص شی برجسته در حوزه‌های متفاوتی مورد استفاده قرار می‌گیرد مانند: تشخیص و شناسایی شی<sup>۱</sup> [۵۴، ۵۳]، فشرده سازی تصویر و ویدئو [۵۶، ۵۵]، خلاصه ویدئو<sup>۲</sup> [۵۷]، کلاژ/هدف گذاری مجدد/برش/تصویر کوچک عکس<sup>۳</sup> [۵۸]، ارزیابی کیفیت تصویر [۵۹]، تقسیم بندی تصویر [۶۰]، بازیابی تصویر مبتنی بر محتوا و مرور مجموعه تصاویر<sup>۴</sup> [۶۱]، ویرایش و دستکاری تصویر [۶۲]، ردیابی بصری<sup>۵</sup> [۶۳]، کشف شی [۶۴] و تعامل انسان و ربات [۶۵]. شکل ۱ تعدادی از کاربردهای تشخیص شی برجسته را نشان می‌دهد.

---

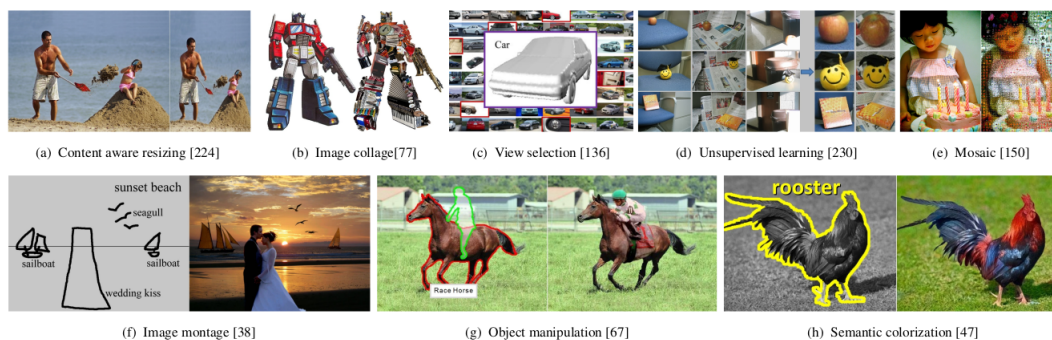
۱ Object detection and recognition

۲ Video summarization

۳ Photo collage/retargeting/cropping/thumbnailing

۴ Content-based image retrieval and image collection browsing

۵ Visual tracking



شکل ۱. تعدادی از کاربردهای تشخیص شی برجسته

## ۴ مجموعه داده‌ها و معیارهای ارزیابی

مجموعه داده‌ها و معیارهای ارزیابی که در کارهای تشخیص شی برجسته استفاده شده‌اند در این بخش نام برده می‌شوند.

### ۱.۴ معیارهای ارزیابی

معیارهای ارزیابی که در این کارها استفاده شده‌اند عبارت‌اند از: Precision-recall (PR), F-measure, Receiver operating characteristics (ROC) curve, Area under ROC curve (AUC), Mean absolute error (MAE).

### ۲.۴ مجموعه داده‌ها

مجموعه داده‌هایی که برای تشخیص شی برجسته استفاده شده‌اند عبارت‌اند از: MSRA-A, MSRA-B, SED1, SED2, ASD, SOD, iCoSeg, MSRA5K, Infrared, ImgSal, CSSD, ECSSD, MSRA10K, THUR15K, DUT-OMRON, Bruce-A, Judd-A, PASCAL-S, UCSB, OSIE, RSD, STC.



- [1] Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba, “Mit saliency benchmark,” 2015. [http://saliency.mit.edu/results\\_mit300.html](http://saliency.mit.edu/results_mit300.html). 1
- [2] Z. Bylinskii, A. Recasens, A. Borji, A. Oliva, A. Torralba, and F. Durand, “Where should saliency models look next?,” in *Computer Vision – ECCV 2016* (B. Leibe, J. Matas, N. Sebe, and M. Welling, eds.), (Cham), pp. 809–824, Springer International Publishing, 2016. [https://doi.org/10.1007/978-3-319-46454-1\\_49](https://doi.org/10.1007/978-3-319-46454-1_49). 1
- [3] A. C. Berg, T. L. Berg, H. Daumé, J. Dodge, A. Goyal, X. Han, A. Mensch, M. Mitchell, A. Sood, K. Stratos, and K. Yamaguchi, “Understanding and predicting importance in images,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3562–3569, 2012. <https://doi.org/10.1109/CVPR.2012.6248100>. 1
- [4] B. M. ’t Hart, H. Schmidt, C. Roth, and W. Einhauser, “Fixations on objects in natural scenes: dissociating importance from salience,” *Frontiers in Psychology*, vol. 4, 2013. <https://doi.org/10.3389/fpsyg.2013.00455>. 1
- [5] M. Spain and P. Perona, “Measuring and predicting object importance,” *International Journal of Computer Vision*, vol. 91, pp. 59–76, Jan 2011. <https://doi.org/10.1007/s11263-010-0376-0>. 1
- [6] P. Isola, J. Xiao, A. Torralba, and A. Oliva, “What makes an image memorable?,” in *CVPR 2011*, pp. 145–152, 2011. <https://doi.org/10.1109/CVPR.2011.5995721>. 1
- [7] R. Rosenholtz, Y. Li, and L. Nakano, “Measuring visual clutter,” *Journal of Vision*, vol. 7, pp. 17–17, 08 2007. <https://doi.org/10.1167/7.2.17>. 1
- [8] S. Dhar, V. Ordonez, and T. L. Berg, “High level describable attributes for predicting aesthetics and interestingness,” in *CVPR 2011*, pp. 1657–1664, 2011. <https://doi.org/10.1109/CVPR.2011.5995467>. 1
- [9] M. Gygli, H. Grabner, H. Riemenschneider, F. Nater, and L. V. Gool, “The interestingness of images,” in *2013 IEEE International Conference on Computer Vision*, pp. 1633–1640, 2013. <https://doi.org/10.1109/ICCV.2013.205>. 1
- [10] Y.-G. Jiang, Y. Wang, R. Feng, X. Xue, Y. Zheng, and H. Yang, “Understanding and predicting interestingness of videos,” in *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, AAAI’13, p. 1113–1119, AAAI Press, 2013. <https://dl.acm.org/doi/10.5555/2891460.2891615>. 1

- [11] H. Katti, K. Y. Bin, T. S. Chua, and M. Kankanhalli, “Pre-attentive discrimination of interestingness in images,” in *2008 IEEE International Conference on Multimedia and Expo*, pp. 1433–1436, 2008. <https://doi.org/10.1109/ICME.2008.4607714>. 1
- [12] L. Itti and P. Baldi, “Bayesian surprise attracts human attention,” in *Advances in Neural Information Processing Systems* (Y. Weiss, B. Schölkopf, and J. Platt, eds.), vol. 18, MIT Press, 2005. <https://papers.nips.cc/paper/2005/hash/0172d289da48c48de8c5ebf3de9f7ee1-Abstract.html>. 1
- [13] Z. Wang, A. C. Bovik, and L. Lu, “Why is image quality assessment so difficult?,” in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. IV–3313–IV–3316, 2002. <https://doi.org/10.1109/ICASSP.2002.5745362>. 1
- [14] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. <https://doi.org/10.1109/TIP.2003.819861>. 1
- [15] W. Zhang, A. Borji, Z. Wang, P. Le Callet, and H. Liu, “The application of visual saliency models in objective image quality assessment: A statistical evaluation,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1266–1278, 2016. <https://doi.org/10.1109/TNNLS.2015.2461603>. 1
- [16] K. A. Ehinger, J. Xiao, A. Torralba, and A. Oliva, “Estimating scene typicality from human ratings and image features,” *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 33, 2011. <https://escholarship.org/uc/item/19n7p532>. 1
- [17] J. Vogel and B. Schiele, “A semantic typicality measure for natural scene categorization,” in *Pattern Recognition* (C. E. Rasmussen, H. H. Bülthoff, B. Schölkopf, and M. A. Giese, eds.), (Berlin, Heidelberg), pp. 195–203, Springer Berlin Heidelberg, 2004. [https://doi.org/10.1007/978-3-540-28649-3\\_24](https://doi.org/10.1007/978-3-540-28649-3_24). 1
- [18] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth, “Describing objects by their attributes,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1778–1785, 2009. <https://doi.org/10.1109/CVPR.2009.5206772>. 1
- [19] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998. <https://doi.org/10.1109/34.730558>. 1, 2

- [20] T. Liu, J. Sun, N.-N. Zheng, X. Tang, and H.-Y. Shum, “Learning to detect a salient object,” in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007. <https://doi.org/10.1109/CVPR.2007.383047>. 1
- [21] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, “Learning to detect a salient object,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011. <https://doi.org/10.1109/TPAMI.2010.70>. 1, 2
- [22] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, “Salient region detection and segmentation,” in *Computer Vision Systems* (A. Gasteratos, M. Vincze, and J. K. Tsotsos, eds.), (Berlin, Heidelberg), pp. 66–75, Springer Berlin Heidelberg, 2008. [https://doi.org/10.1007/978-3-540-79547-6\\_7](https://doi.org/10.1007/978-3-540-79547-6_7). 1, 2
- [23] Y.-F. Ma and H.-J. Zhang, “Contrast-based image attention analysis by using fuzzy growing,” in *Proceedings of the Eleventh ACM International Conference on Multimedia*, MULTIMEDIA ’03, (New York, NY, USA), p. 374–381, Association for Computing Machinery, 2003. <https://doi.org/10.1145/957013.957094>. 1
- [24] F. Liu and M. Gleicher, “Region enhanced scale-invariant saliency detection,” in *2006 IEEE International Conference on Multimedia and Expo*, pp. 1477–1480, 2006. <https://doi.org/10.1109/ICME.2006.262821>. 1, 2, 3
- [25] D. Walther and C. Koch, “Modeling attention to salient proto-objects,” *Neural Networks*, vol. 19, no. 9, pp. 1395–1407, 2006. Brain and Attention. 1
- [26] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. <https://doi.org/10.1109/5.726791>. 2
- [27] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015. <https://doi.org/10.1109/CVPR.2015.7298965>. 2
- [28] Y. Hu, D. Rajan, and L.-T. Chia, “Robust subspace analysis for detecting visual attention regions in images,” in *Proceedings of the 13th Annual ACM International Conference on Multimedia*, MULTIMEDIA ’05, (New York, NY, USA), p. 716–724, Association for Computing Machinery, 2005. <https://doi.org/10.1145/1101149.1101306>. 2
- [29] P. L. Rosin, “A simple method for detecting salient regions,” *Pattern Recognition*, vol. 42, no. 11, pp. 2363–2371, 2009. <https://doi.org/10.1016/j.patcog.2009.04.021>. 2

- [30] R. Valenti, N. Sebe, and T. Gevers, “Image saliency by isocentric curvedness and color,” in *2009 IEEE 12th International Conference on Computer Vision*, pp. 2185–2192, 2009. <https://doi.org/10.1109/ICCV.2009.5459240>. 2
- [31] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, “Frequency-tuned salient region detection,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604, 2009. <https://doi.org/10.1109/CVPR.2009.5206596>. 2
- [32] D. A. Klein and S. Frintrop, “Center-surround divergence of feature statistics for salient object detection,” in *2011 International Conference on Computer Vision*, pp. 2214–2219, 2011. <https://doi.org/10.1109/ICCV.2011.6126499>. 2
- [33] X. Li, Y. Li, C. Shen, A. Dick, and A. V. D. Hengel, “Contextual hypergraph modeling for salient object detection,” in *2013 IEEE International Conference on Computer Vision*, pp. 3328–3335, 2013. <https://doi.org/10.1109/ICCV.2013.413>. 2
- [34] S. Goferman, L. Zelnik-Manor, and A. Tal, “Context-aware saliency detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012. <https://doi.org/10.1109/TPAMI.2011.272>. 2
- [35] R. Margolin, A. Tal, and L. Zelnik-Manor, “What makes a patch distinct?,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1139–1146, 2013. <https://doi.org/10.1109/CVPR.2013.151>. 2, 3
- [36] Z. Yu and H.-S. Wong, “A rule based technique for extraction of visual attention regions based on real-time clustering,” *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 766–784, 2007. <https://doi.org/10.1109/TMM.2007.893351>. 3
- [37] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, “Global contrast based salient region detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015. <https://doi.org/10.1109/TPAMI.2014.2345401>. 3
- [38] Z. Jiang and L. S. Davis, “Submodular salient region detection,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2043–2050, 2013. <https://doi.org/10.1109/CVPR.2013.266>. 3
- [39] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, “Saliency filters: Contrast based filtering for salient region detection,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 733–740, 2012. <https://doi.org/10.1109/CVPR.2012.6247743>. 3

- [40] C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, and D. A. Clausi, “Statistical textural distinctiveness for salient region detection in natural images,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 979–986, 2013. <https://doi.org/10.1109/CVPR.2013.131>. 3
- [41] K. Shi, K. Wang, J. Lu, and L. Lin, “Pisa: Pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2115–2122, 2013. <https://doi.org/10.1109/CVPR.2013.275>. 3
- [42] H. Li and K. N. Ngan, “A co-saliency model of image pairs,” *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3365–3375, 2011. <https://doi.org/10.1109/TIP.2011.2156803>. 3
- [43] L. Marchesotti, C. Cifarelli, and G. Csurka, “A framework for visual saliency detection with applications to image thumbnailing,” in *2009 IEEE 12th International Conference on Computer Vision*, pp. 2232–2239, 2009. <https://doi.org/10.1109/ICCV.2009.5459467>. 3
- [44] H. Fu, X. Cao, and Z. Tu, “Cluster-based co-saliency detection,” *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 3766–3778, 2013. <https://doi.org/10.1109/TIP.2013.2260166>. 3
- [45] S. He, R. W. H. Lau, W. Liu, Z. Huang, and Q. Yang, “Supercnn: A superpixelwise convolutional neural network for salient object detection,” *International Journal of Computer Vision*, vol. 115, pp. 330–344, Dec 2015. <https://doi.org/10.1007/s11263-015-0822-0>. 3
- [46] L. Wang, H. Lu, X. Ruan, and M.-H. Yang, “Deep networks for saliency detection via local estimation and global search,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3183–3192, 2015. <https://doi.org/10.1109/CVPR.2015.7298938>. 3
- [47] R. Zhao, W. Ouyang, H. Li, and X. Wang, “Saliency detection by multi-context deep learning,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1265–1274, 2015. <https://doi.org/10.1109/CVPR.2015.7298731>. 3
- [48] G. Li and Y. Yu, “Visual saliency based on multiscale deep features,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5455–5463, 2015. <https://doi.org/10.1109/CVPR.2015.7299184>. 3
- [49] N. Liu and J. Han, “Dhsnet: Deep hierarchical saliency network for salient object detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 678–686, 2016. <https://doi.org/10.1109/CVPR.2016.80>. 3

- [50] G. Li and Y. Yu, “Deep contrast learning for salient object detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 478–487, 2016. <https://doi.org/10.1109/CVPR.2016.58>. 3
- [51] J. Kuen, Z. Wang, and G. Wang, “Recurrent attentional networks for saliency detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3668–3677, 2016. <https://doi.org/10.1109/CVPR.2016.399>. 3
- [52] S. S. S. Kruthiventi, V. Gudisa, J. H. Dholakiya, and R. V. Babu, “Saliency unified: A deep architecture for simultaneous eye fixation prediction and salient object segmentation,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5781–5790, 2016. <https://doi.org/10.1109/CVPR.2016.623>. 3
- [53] A. Borji, M. N. Ahmadabadi, and B. N. Araabi, “Cost-sensitive learning of top-down modulation for attentional control,” *Machine Vision and Applications*, vol. 22, pp. 61–76, Jan 2011. <https://doi.org/10.1007/s00138-009-0192-0>. 3
- [54] H. Shen, S. Li, C. Zhu, H. Chang, and J. Zhang, “Moving object detection in aerial video based on spatiotemporal saliency,” *Chinese Journal of Aeronautics*, vol. 26, no. 5, pp. 1211–1217, 2013. <https://doi.org/10.1016/j.cja.2013.07.038>. 3
- [55] C. Guo and L. Zhang, “A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression,” *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 185–198, 2010. <https://doi.org/10.1109/TIP.2009.2030969>. 3
- [56] L. Itti, “Automatic foveation for video compression using a neurobiological model of visual attention,” *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1304–1318, 2004. <https://doi.org/10.1109/TIP.2004.834657>. 3
- [57] Q.-G. Ji, Z.-D. Fang, Z.-H. Xie, and Z.-M. Lu, “Video abstraction based on the visual attention model and online clustering,” *Signal Processing: Image Communication*, vol. 28, no. 3, pp. 241–253, 2013. <https://doi.org/10.1016/j.image.2012.11.008>. 3
- [58] S. Goferman, A. Tal, and L. Zelnik-Manor, “Puzzle-like collage,” *Computer Graphics Forum*, vol. 29, no. 2, pp. 459–468, 2010. <https://doi.org/10.1111/j.1467-8659.2009.01615.x>. 3
- [59] A. Li, X. She, and Q. Sun, “Color image quality assessment combining saliency and FSIM,” in *Fifth International Conference on Digital Image Processing (ICDIP 2013)* (Y. Wang and X. Yi, eds.), vol. 8878,

- p. 88780I, International Society for Optics and Photonics, SPIE, 2013. <https://doi.org/10.1109/CVPR.2012.6248100>. 3
- [60] M. Donoser, M. Urschler, M. Hirzer, and H. Bischof, “Saliency driven total variation segmentation,” in *2009 IEEE 12th International Conference on Computer Vision*, pp. 817–824, 2009. <https://doi.org/10.1109/ICCV.2009.5459296>. 3
- [61] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, “Sketch2photo: Internet image montage,” *ACM Trans. Graph.*, vol. 28, p. 1–10, dec 2009. <https://doi.org/10.1145/1618452.1618470>. 3
- [62] A. Y.-S. Chia, S. Zhuo, R. K. Gupta, Y.-W. Tai, S.-Y. Cho, P. Tan, and S. Lin, “Semantic colorization with internet images,” *ACM Trans. Graph.*, vol. 30, p. 1–8, dec 2011. <https://doi.org/10.1145/2070781.2024190>. 3
- [63] A. Borji, S. Frintrop, D. N. Sihite, and L. Itti, “Adaptive object tracking by learning background context,” in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 23–30, 2012. <https://doi.org/10.1109/CVPRW.2012.6239191>. 3
- [64] S. Frintrop, G. M. García, and A. B. Cremers, “A cognitive approach for object discovery,” in *2014 22nd International Conference on Pattern Recognition*, pp. 2329–2334, 2014. <https://doi.org/10.1109/ICPR.2014.404>. 3
- [65] D. Meger, P.-E. Forssén, K. Lai, S. Helmer, S. McCann, T. Southey, M. Baumann, J. J. Little, and D. G. Lowe, “Curious george: An attentive semantic robot,” *Robotics and Autonomous Systems*, vol. 56, no. 6, pp. 503–511, 2008. From Sensors to Human Spatial Concepts. 3