

# The Art of Questioning Lethal Vision: Mosse's *Infra* and Militarized Machine Vision

Rune Saugmann  
International Politics  
University of Tampere  
Tampere, Finland  
[rune.saugmann@uta.fi](mailto:rune.saugmann@uta.fi)

**Military techno-vision is decidedly opaque. We – the public in whose name they are developed and deployed – are never quite allowed to scrutinize how exactly it operates, or what kind of work it performs. For a long time, military techno-vision was related to the optical qualities of the human eye. It was either about enhancing the natural human eye, seeing further, seeing in less light, seeing from different locations, or about mimicking the human eye, creating spatial awareness and navigation capabilities from the singular point of the eye/camera. Today, increasingly, militarized techno-vision is not merely about the production but also the interpretation of images, about the mind rather than the eye, to use the anthropomorphic language that dominates high-tech discourse. This shift from optics to procession, vision or understanding is related to the development of precision optics and the resulting flood of images that threatens to overwhelm the security bureaucracies seeking to optimize lethal vision. The contemporary problem, thus, is not overcoming invisibility but rather how to efficiently manage a flood of opaque images.**

**The solution, proposed by militaries as simultaneously benign and efficient, is to delegate image interpretation to machines. In this paper, I ask how we can scrutinize and interrogate the agency of such machine vision systems when we as a public are rarely granted access to the operations they perform. First, I briefly explore current machine vision technologies, and the technological projects that have identified their blind spots. Second, I look to artistic image production as a means to help think about militarized vision and pictorial agency.**

*Machine vision. Art Photography. Security.*

## 1. INTRODUCTION

Military precision optics and, increasingly, automated forms of military techno-vision play crucial roles in the assemblages that manage and make sense of global politics, making sense of sociality and inscribing categorizations and lethal decisions in different quotidian realities of the 'everywhere war' (Gregory, 2011; Tazzioli and Walters, 2016; Walters, 2014). In this paper, I ask

how we can make sense of the sense-making activities of military techno-vision systems that we are not given access to observe. I argue that while western publics are asked to place their trust in these systems, a cloak of invisibility shields military techno-optics from civilian oversight or insight – attempts to see through or with them. In contrast to the superhuman vision they allegedly provide, the systems are kept opaque and inaccessible for the western publics who are not allowed to peer through

their scopes or to scrutinize the code that they implement onto the world.

In this exploratory short paper, I ask how we can scrutinize and interrogate the agency of such machine vision systems when we as a public are rarely granted access to the operations they perform.

Critical understanding can depart from the videos made publicly available – predominantly promotional videos published by western militaries applying military techno-vision in war, but the semiotic techniques applied in such promotional material serves to mask the confusion of the battlefield and render alternative understandings unlikely. Likewise, even if the 'collateral murder' video leaked by Chelsea Manning – by far the most prominent actual documentation of the air-war practices developed under the continuations of the war on terror - provided a strong indictment of the practices of warfare guided by precision optics, I have argued elsewhere that authorities were able to largely brush this critique aside by deploying what I call a 'semiotic fog of war' that disenables their visual critique (Andersen, 2015).

To make sense of the sense-making done in and by military techno-vision, we also have to look elsewhere, to decidedly non-documentarist images. In Richard Mosse's artworks 'the enclave', military film designed to reveal what is invisible to the unassisted eye is used to photograph a little-known war, decades-long conflict in eastern Congo. Here, western spectators do not easily identify and classify 'good' and 'bad' actors, forcing us to think differently. By rendering the conflict a beautiful and surreal mess that is hard to read 'operationally', the photographs serve as an invitation for us to see the interpretative work done through but also left to soldiers by technological enhanced forms of military vision. In the following, I first briefly explore the infrastructuring of current machine vision technologies in relation to security, and some of the technological projects that have identified their blind spots. Arguing that the degree to which we can make sense of such systems based on this kind of knowledge is not fully satisfactory, I, second, look to Mosse's artistic image production as a means to help think about militarized vision and pictorial agency.

## **2. MACHINES, VISION, AND SECURITY**

After decades of development of ever-more powerful surveillance technology, the problem of visibility for security services is no longer sensing or optics, it is vision (or understanding). Terrorism attacks such as the 2012 Breivik leave the authorities with tens of thousands of hours of relevant surveillance footage to review (Norwegian Police Security Service (PST), g). Half a decade

ago, in 2011, 'the US Air Force [had] amassed over 325,000 hours of drone video—that's about 37 years of video' and today a 'single drone with these sensors produces many terabytes of data every day. Before AI was incorporated into analysis of this data, it took a team of analysts working 24 hours a day to exploit only a fraction of one drone's sensor data' (Allen, 2017). These are not only data problems, as already in 2010 the fog of war had become digital in the sense that the US had to blame drone strikes targeting civilians on 'information overload' (Johnson and Wald, 2017). Currently, the situation is thus that 'the defense intelligence community is currently drowning in data. Every day, US spy planes and satellites collect more raw data than the Defense Department could analyze even if its whole workforce spent their entire lives on it' (Allen, 2017)

Machine vision, the development of neural network technology to categorise content in images and search for patterns in them or for matches with existing database is the current proposition for turning observation into the militarised version of understanding, actionable intelligence. Not being an engineer, I entered the field of machine vision software through controversies that are accessible to outsiders, starting from the training and testing image databases and from there working my way towards machine vision systems that are trained through them and the visual culture they inherit from such databases.

### **2.1 Training databases, ambiguity and ground truth**

The most widely used training image database, ImageNet is built by having search engines look for images following simple descriptors in major search engines, i.e. consists of images harvested online (Deng et al., 2009, p. 251). Search engine images, of course, are not entirely unbiased, and the constructors of ImageNet remark that '[t]o further enlarge and diversify the candidate pool, we translate the queries into other languages, including Chinese, Spanish, Dutch and Italian' (Deng et al., 2009, p. 251).

Images are described and categorized using the labour-sourcing portal Mechanical Turk, where workers get paid in fractions of cents for each very small job done, to produce and verify descriptions before entering the image into the database which in 2014 'contain[ed] 14,197,122 annotated images organized by the semantic hierarchy of WordNet' (Russakovsky et al., 2014, pp. 7, 3). The database contains more than twenty thousand categories of objects that machine vision systems can then be trained to recognize. The database hosts an annual machine vision competition in which systems compete in *image classification* - which 'tests the ability of an algorithm to name the objects present in the image, without necessarily localizing them';

*object detection and localization* – which ‘evaluates the ability of an algorithm to localize one instance of an object category; and *object detection* – which evaluates the ability of an algorithm to name and localize all instances of all target objects present in an image (Russakovsky et al., 2014, pp. 5, 8, 10). The competition is based on images harvested from Flickr (Russakovsky et al., 2014, p. 4) and has been credited with contributing to the recent acceleration in advances in deep learning following a breakthrough made in the 2012 contest (The Economist, 2016).

A later database, Microsoft Common Objects in Context (MS COCO), is constructed using Flickr images against the idea that ‘current recognition systems perform fairly well on iconic views, but struggle to recognize objects otherwise – in the background, partially occluded, amid clutter – reflecting the composition of actual everyday scenes.’ (Lin et al., 2015, p. 1). This is the result of ImageNet (and other datasets’) images being sourced primarily through search engine querying, producing ‘iconic’ images in which the sought-after object or attribute is paramount, leading to images that according to the MS COCO developers ‘lack important contextual information and non-canonical viewpoints’ (Lin et al., 2015, p. 4). In such settings, ‘Object category presence is often ambiguous. Indeed ... even dedicated experts often disagree on object presence, e.g. due to inherent ambiguity in the image or disagreement about category definitions (Lin et al., 2015, p. 6). Yet, as machine teaching goes, the dataset still works by assigning a ‘ground truth’, ‘computed using majority vote of the experts’ (Lin et al., 2015, p. 6) to denote what the image ‘truly’ contains.

Several features of these datasets are interesting. One is about what kind of use people’s images are put to without their knowing – which is perhaps not so intuitively alarming when developing machine vision in general but attain a different ethical depth once machine vision is combined with weapons or surveillance systems. Even if freely available online, should unknown people’s images been subjected to this kind of use without their consent? And by relying heavily on Flickr and image search engines, machine vision system learns from and within the specific visual culture that is Flickr or that is already monetised by search engines. A rather funny illustration of this from machine vision is from an allegory told in an Economist editorial, about the ‘unsupervised’ discovery of classes of objects: ‘Google Brain in which a giant unsupervised learning system was asked to look for common patterns in thousands of unlabelled YouTube videos. One day one of Mr Ng’s PhD students had a surprise for him. “I remember him calling me over to his computer and saying, ‘look at this’,” Mr Ng recalls. On the screen

was a furry face, a pattern distilled from thousands of examples. The system had discovered cats.’ (The Economist, 2016). Anybody familiar with YouTube and the output of its recommendation algorithms (Saugmann Andersen, 2017) will know that cats are at the apex of an enormous video-industrial complex, making the discovery of cats also a feedback loop discovering algorithmic governance, rather than merely furry animals.

Given the importance of and difficulty with datasets for developing machine vision systems, these will likely also be an issue for surveillance and military applications of machine vision. Training datasets are thus, unsurprisingly, an issue in publicly known military applications of machine vision. The most well-known, Project Maven, a US Department of Defense project to establish an Algorithmic Warfare Cross-Function Team, was widely reported in 2018 due to the revelation that Google was involved despite promises to not weaponise artificial intelligence, and the subsequent protests by Google employees.

The project’s ‘objective is to turn the enormous volume of data available to the DoD into actionable intelligence and insights at speed’, with the project’s first task to ‘organize a data-labeling effort, and develop, acquire, and/or modify algorithms to accomplish key tasks’, these being to ‘reduce the human factors burden [i.e. labour needs] of FMV [Full-Motion Video] analysis, increase actionable intelligence, and enhance military decision-making’ (Deputy Secretary of Defense, 2017). Project Maven thus ‘seeks to automate basic labeling and analysis associated with full-motion video surveillance’ (Johnson and Wald, 2017) and ‘focuses on analysis of full-motion video data from tactical aerial drone platforms such as the ScanEagle and medium-altitude platforms such as the MQ-1C Gray Eagle and the MQ-9 Reaper.’ (Allen, 2017), some of the drones responsible for the visual data deluge.

Even if Maven focuses on one class of images only, the differences in training data are striking.

‘In Maven’s case, humans had to individually label more than 150,000 images in order to establish the first training data sets; the group hopes to have 1 million images in the training data set by the end of January [2018] (Allen, 2017). Whether one or the other – noting that hopes and future promises in military technology are unlikely to describe any reality – the paucity of training images is striking compared to the lethal applications the software is aimed at, and the way in which ImageNet’s 14 million images were described above as mainly giving ‘iconic’ views of the objects identified in them. The approach to deal with the lack of labelled data seems to be one of limiting the capabilities of the system – or, if we adopt the anthropomorphic lingo of artificial intelligence and machine vision, one of visual stupidity. Maven’s ‘immediate focus is 38 classes of objects that represent the kinds of things

the department needs to detect, especially in the fight against the Islamic State of Iraq and Syria.' (Pellerin, 2017). While it may be desirable to start out with a narrow focus, such a focus obviously doesn't alter the ambiguity and multiplicity that is the 'ground truth' lived on the ground, it only sharpens the reductionist elements of the machinic 'ground truth'.

Militarized machine vision will need training datasets that not only permits the systems learning from them to segment objects and distinguish between them but also be highly sensitive to context (think of a fighter with a gun versus a farmer with a gun) and visual culture (e.g. similar gestures with different cultural meanings).

## 2.2 Machine misrecognition

Seeking to get an idea of how machine vision works, and which controversies are debated within the community developing it, I eventually arrived at Google's publicly available image description and object identification systems. These, I thought, would have to stand in for the secret but probably similar systems that I suspected would be at worked in military image databases, and this section was written on that background.

Yet the connection between in the first months of 2018, it was revealed that Google was indeed providing machine vision services to the US Department of Defense, through Project Maven. The descriptions of this project as an effort that 'will 'provide computer vision algorithms for object detection, classification, and alerts for FMV [full-motion video] PED [processing, exploitation, and dissemination]' (Deputy Secretary of Defense, 2017) brought uncanny memories of the features I had laughed at in descriptions of the image description and object labelling system.

Google's image description and object identification system works through combining different neural networks, as shown in figure 1, combining machine vision with natural language processing.

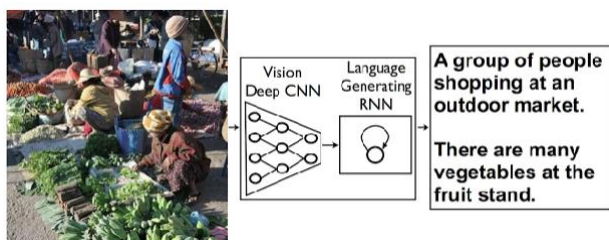


Figure 1: Figure 2: Representation of Google's image description and object identification system (Vinyals et al., 2015, p. 1)

The use value of this combination lies partly in its making images compatible with other database

objects, and easily searchable through the keywords and object descriptors generated. This epistemic rendering of the image rests on, firstly, an understanding of images as transparent or non-distorting – what can be termed an aesthetic of immediacy (cf. Bolter and Grusin, 2000). Secondly, it rests on an understanding of reality as limited and un-ambiguous. These properties are striking in Google's description of its machine vision system:

*“Two pizzas sitting on top of a stove top oven” “A group of people shopping at an outdoor market” “Best seats in the house”* People can summarize a complex scene in a few words without thinking twice. It's much more difficult for computers. But we've just gotten a bit closer -- we've developed a machine-learning system that can automatically produce captions (like the three above) to accurately describe images the first time it sees them.' (Vinyals et al., 2014)

The claims above are strikingly at odds with the understanding of images that one would get from visual studies, visual culture studies, visual semiotics or the like – and the sentences do not sound like how most people would describe any scene or image, yet the creators assure that the system meets state-of-the-art thresholds and comes close to 'human performance' (Vinyals et al., 2015) even with the added difficulty of formulating image captions in sentences. The problem that concerns us at the moment, though, is not whether or not images can be reduced to ground truths, but rather the technological fragility of this translation.

In an intriguing article titled 'Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images' (Nguyen et al., 2015), the authors show that 'it is easy to produce images that are completely unrecognizable to humans, but that state-of-the-art DNNs believe to be recognizable objects with 99.99% confidence'. Using machine vision systems trained on ImageNet or similar large databases, and pairing these with evolutionary algorithms 'that optimize images to generate high-confidence DNN predictions for each class in the dataset the DNN is trained on' (Nguyen et al., 2015). The technical process for doing so is not my concern here, rather the character of the images and their relation to the type of understanding sought.



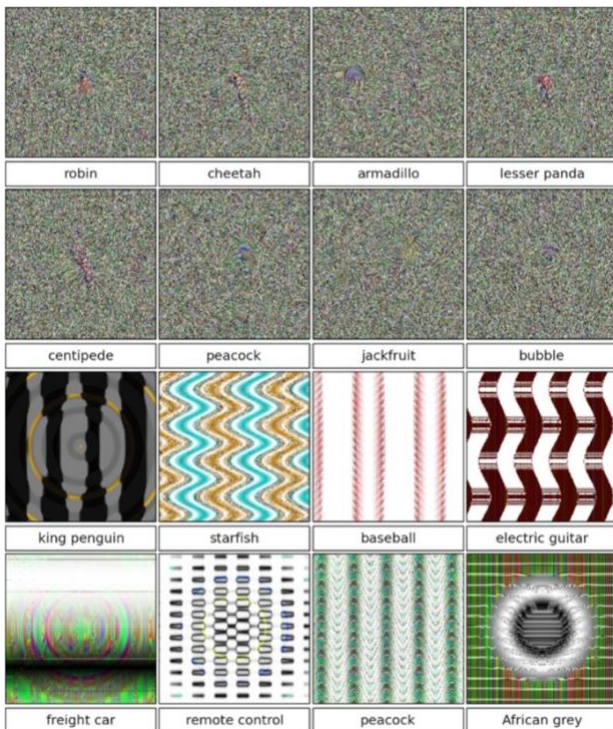


Figure 2: Fooling images as illustrated by Nguyen et al. (2015)

The strangeness of these images is intriguing, funny, and of course deeply uncanny when thinking of how Maven and similar military projects *already* use machine vision to classify objects of concern in the battlefield (Allen, 2017). While the images in the lower panes betray some abstract symmetry and perhaps likeness to that which they are mistaken as, the images above show clearly that whatever machine understanding of images is, it is not vision.

Thus while machine vision systems may perform well in terms of ‘actionable intelligence’, the gold standard of military superiority, the form of intelligence is fundamentally unrelated to what we call ‘vision’ in normal human interaction. Perhaps this is related to the introduction of engineering ideas such as ‘ground truth’ into the domain of changing social conventions that we call vision.

Surprisingly, the errors uncovered here seem to be rather robust, as ‘it is not easy to prevent the DNNs from being fooled by retraining them with fooling images labeled as such. While retrained DNNs learn to classify the negative examples as fooling images, a new batch of fooling images can be produced that fool these new networks, even after many retraining iterations.’ (Nguyen et al., 2015, p. 2)

### 3. ART PHOTOGRAPHY AND TECHNOLOGICAL VISION

Art photographer Richard Mosse’s *Infra* photo series and its video component *the enclave* uses Kodak Aerochrome, a discontinued military reconnaissance film, to portray scenes from eastern Congo. The film has an extra infrared exposure layer which enables it to register chlorophyll in live vegetation, and was used during the Korea and Viet Nam wars for aerial vegetation surveillance, and deployed to reveal camouflage. According to the producer it is ‘an infrared-sensitive, false-color reversal film intended for various aerial photographic applications where infrared discriminations may yield *practical results*’ (Kodak, 2005, my italics). The way in which the Aerochrome film produces these practical results is by colouring plants with living/active chlorophyll – the agent of photosynthesis – pink; while leaving the dead organic or non-organic material used in camouflaging efforts in its original colours. The result is beautiful images rendered with an array of shades of pink, landscapes and portraits of beauty but also of suspicion.

The film was at its time a way of military vision, ‘developed by the US military in the 1940s to detect camouflage and to reveal part of the spectrum of light the human eye cannot see.’ (Stearns, 2011) thus permitting the military using it to see through the deception strategies deployed by adversaries.

As explained by the producer, ‘[c]olor infrared-sensitive films were originally designed for reconnaissance and camouflage detection. In fact, the term “CD” was once used to denote the camouflage detection role of this film.

In *Infra*, Mosse uses Aerochrome to take pictures of scenes in Eastern Congo, in areas torn by a brutal and decades-long civil war that for western spectators offers no easy distinctions between ‘good’ and ‘evil’. This visual strangeness of the conflict is important as ‘the sense derived from seen objects is not merely a function of the degree of optical resolution; it derives from the projects and culturally induced expectations of the observer’ (Shapiro, 2009, p. 66). With soldiers and militants in Congo lacking markers that would distinguish them visually as friend and enemy, perpetrator or protector, Mosse achieves a distancing of the western spectator that perhaps can serve to make this spectator sensible to the agency of the sensing technology.

In *Infra*, the erratic colouring becomes dizzying rather than legible, a destabilizing effect of the pink ‘magic’ of showing the world unknown to us in terms of foliage and deception, a beautiful constraint on intelligibility rather than actionable military intelligence. The Aerochrome film makes everyday images surreal – beautiful, perhaps, but also

confusing. To Mosse, this misfit was what stirred his interest in using the film in Congo, remarking that 'it seemed inappropriate and made me feel slightly uncomfortable' (Rebibo, 2014).

Crucially, Mosse's Aerochrome images are opaque and hard to read operationally, their distribution of real and unreal unreliable and confusing rather than readily providing the 'actionable intelligence' touted by military visualization technology producers and users. It reveals a world of technologically enhanced vision that questions its own applicability. Soldiers are marked as living plant materials, civilian housing stands out marked as dangerous, landscapes full of suspicion meet us and there appears to be at least as many questions after the use of the film as there were before. Where the filmic technology promises to extend our bodily abilities of seeing to leave us with a clearer view of opaque battlefields and 'detect enemy positions in the underbrush' (Stearns, 2011), Mosse exposes the visual fragility of such a technology that is extremely fragile both in representative and practical ways – it must be kept cold until exposure, and only very few studios can develop it.

Most observers have called attention to the way in which the resulting photography plays with tropes of Congo and Africa as the 'dark continent', and truly it does. But it is also able to 'make us call into question pictures we thought we understood.' (Stearns, 2011) and, I think, asks us to question their apparatuses of distinction they are part of (Perugini and Gordon, 2017), the militarized technologies we are being asked to believe in but which we are not given access to scrutinize.

#### 4. CONCLUSION

The lethal ways of seeing enacted in militarized visual sensing systems depend on our trust in the visual technologies as ways of knowing the social and natural world. This trust, then, forms the basis for the enactment of a 'sovereign gaze' that can support security decisions by producing actionable intelligence (Amoore, 2007). Technological controversies reveal the utopianism underpinning even modest military machine vision projects, such as project Maven's effort towards autonomous labelling of 38 classes of objects, as these rely on much smaller training databases than their impressive but also fragile civilian counterparts. Despite their virtues – including superior databases and international competitions to spur development – these systems are prone to not only systematic misrecognition but to systematic misrecognition of fundamentally non-iconic images as objects. Decades ago, Haraway pointed to the wider agency of devices of vision, remarking how "the "eyes" made available in modern technological sciences shatter any idea of passive vision; these prosthetic devices

show us that all eyes, including our own organic ones, are active perceptual systems, building in translations and specific ways of seeing' (Haraway, 1991, p. 190). What the scatter peacocks misrecognized by machine vision tell about, is not only an active perceptual system but a fundamentally non-human one. Whatever this technology does, vision is not an appropriate description. With Richard Mosse's art photography project *Infra*, we can peer through disused military visual prosthetics to scrutinize a warzone as the militarized visual technologies of the past would do. Thus we can appreciate that not only is technological vision transformative, it contains an inescapable and hidden layer of uncertainty, confusion and mess that is being erased in the conversion of seeing – an always subjective, deeply personal and unstable biological-cultural faculty (Lyon, 1994) – into a militarized form of knowing.

#### 5. REFERENCES

- Allen, G.C. (2017) Bringing AI to the fight against ISIS. *Bulletin of the Atomic Scientists*. <https://thebulletin.org/project-maven-brings-ai-fight-against-isis11374> (retrieved 4.5.18)
- Amoore, L. (2007) Vigilant Visualities: The Watchful Politics of the War on Terror. *Security Dialogue*, 38, 215–232.
- Andersen, R.S. (2015) *Remediating Security. A semiotic framework for analyzing how video speaks security*. University of Copenhagen, Department of Political Science, Copenhagen.
- Bolter, J., Grusin, R.A. (2000) *Remediation: understanding new media*, MIT Press, Cambridge, Mass.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, Li Fei-Fei (2009) ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Miami, FL, USA, 248–255.
- Deputy Secretary of Defense (2017) Establishment of an Algorithmic Warfare Cross-Functional Team (Project Maven).
- Gregory, D. (2011) The everywhere war. *The Geographical Journal*, 177, 238–250.
- Haraway, D.J. (1991) *Simians, Cyborgs, and Women. The Reinvention of Nature*. Free Association Books, London.
- Johnson, T., Wald, C.F. (2017) To Handle Its Influx of Drone Footage, Military Should Teach AI to

- Watch TV. WIRED.  
<https://www.wired.com/story/the-military-should-teach-ai-to-watch-drone-footage/> (retrieved 11.6.18)
- Kodak (2005) KODAK AEROCHROME III Infrared Film 1443.  
[https://www.kodak.com/uploadedFiles/Corporate/Industrial\\_Materials.../ti2562.pdf](https://www.kodak.com/uploadedFiles/Corporate/Industrial_Materials.../ti2562.pdf) (retrieved 11.6.18)
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollár, P. (2015) Microsoft COCO: Common Objects in Context. arXiv:1405.0312v3
- Lyon, D. (1994) *The electronic eye: the rise of surveillance society*. University of Minnesota Press, Minneapolis.
- Nguyen, A., Yosinski, J., Clune, J. (2015) Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. *Computer Vision and Pattern Recognition, IEEE*.  
<https://arxiv.org/abs/1412.1897>
- Norwegian Police Security Service (2014). PST and Societal Security. Presentation at Peace Research Institute Oslo, 12.12.14
- Pellerin, C. (2017) Project Maven to Deploy Computer Algorithms to War Zone by Year's End. U.S. DEPARTMENT OF DEFENSE.  
<https://www.defense.gov/News/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/> (retrieved 4.5.18).
- Perugini, N., Gordon, N. (2017) Distinction and the Ethics of Violence: On the Legal Construction of Liminal Subjects and Spaces: Distinction and the Ethics of Violence. *Antipode* 49, 1385–1405.
- Rebibo, R. (2014) I STILL SHOOT FILM. [istillshootfilm.org](http://istillshootfilm.org).  
<http://istillshootfilm.org/post/48700294601/richard-mosse-interview-kodak-aerochrome-in-the> (retrieved 6.27.18).
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L. (2014) ImageNet Large Scale Visual Recognition Challenge. arXiv:1409.0575 [cs].
- Saugmann Andersen, R. (2017) Video, algorithms, and security. *Security Dialogue* 48, 354–372.
- Shapiro, M.J. (2009) *Cinematic geopolitics*. Routledge, London; New York.
- Stearns, J. (2011) Shocking pink. The Guardian. <http://www.theguardian.com/artanddesign/2011/may/28/richard-mosse-infrared-photos-congo> (retrieved 27.6.18)
- Tazzioli, M., Walters, W. (2016) The Sight of Migration: Governmentality, Visibility and Europe's Contested Borders. *Global Society* 30, 445–464.
- The Economist (2016) From not working to neural networking. The Economist.  
<https://www.economist.com/special-report/2016/06/25/from-not-working-to-neural-networking> (retrieved 4.5.18).
- Vinyals, O., Toshev, A., Bengio, S., Erhan, D. (2015) Show and Tell: A Neural Image Caption Generator. arXiv:1411.4555v2.
- Vinyals, O., Toshev, A., Bengio, S., Erhan, D. (2014) A picture is worth a thousand (coherent) words: building a natural description of images. Google AI Blog.  
<https://ai.googleblog.com/2014/11/a-picture-is-worth-thousand-coherent.html> (retrieved 4.5.18)
- Walters, W. (2014) Drone strikes, dingpolitik and beyond: Furthering the debate on materiality and security. *Security Dialogue* 45, 101–118.