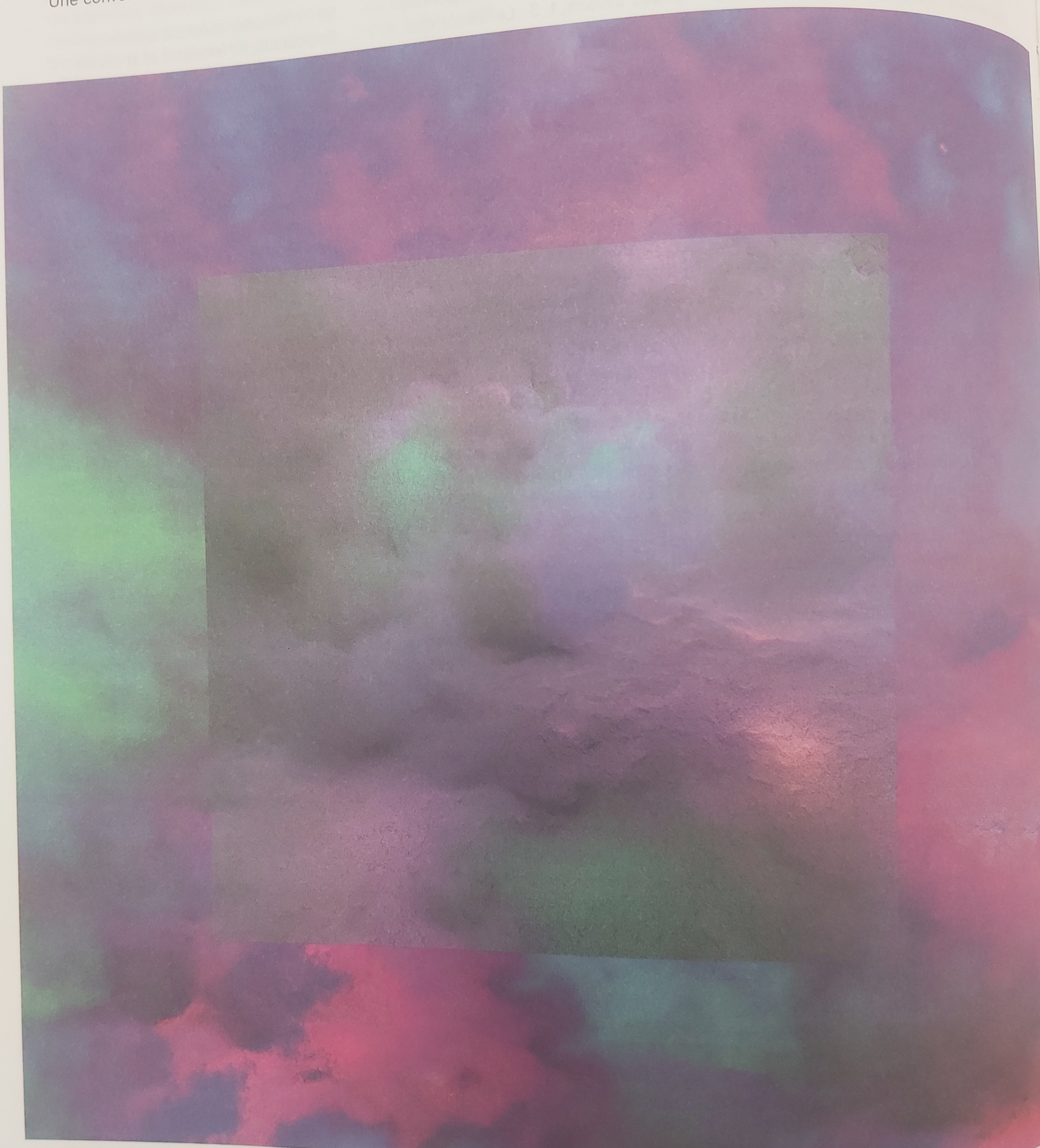
Une conversation entre Grégory Chatonsky, Christian Joschke et Antonio Somaini



1. Grégory Chatonsky, *Preform Latent Space : 37e0dcc969ea32920fe5c8dabaccdd4d*, 2021, bruit visuel généré par un réseau de neurones antérieur à l'apparition de l'image indicielle. Version modifiée de StyleGAN 2.

L'essor récent des intelligences artificielles a transformé radicalement notre rapport au visible en introduisant un nouveau type d'images composites : les images générées automatiquement par des machines. L'artiste Grégory Chatonsky a tiré de ces outils numériques une œuvre foisonnante et étrange, qu'il accompagne d'une réflexion philosophique nourrie de références françaises et anglo-américaines. Le théoricien des images et de la culture visuelle Antonio Somaini, de son côté, pratique une archéologie des médias qui permet de donner au phénomène toute son épaisseur historique. Cette conversation est une amorce de réflexion qui, loin de faire le tour de la question, pourra ouvrir quelques perspectives.

CJ Nous vivons depuis quelques années un tournant majeur dans notre rapport aux images. Avec le développement des intelligences artificielles, nous en voyons apparaître qui prennent l'apparence d'images indicielles – de photographies – mais qui représentent des choses inexistantes car elles sont fabriquées à partir de millions d'images, dont elles produisent une sorte de synthèse plus ou moins aléatoire, guidée par des phrases de commande en langage « naturel » (non informatique) qu'on appelle des *prompts*. Ces images ne renvoient pas à une chose, à un événement, mais sont le résultat d'un calcul et empruntent leurs formes aux images-sources récoltées par millions. Ce sont des images d'images. Ainsi se trouve remis en question le « pacte indiciaire » de la photographie qui nous faisait croire à l'existence d'un référent de la représentation. Certes, cette remise en question de la relation mimétique n'est pas nouvelle. Dès le XIX^e^ siècle, Oscar Rejlander produisait des compositions allégoriques à partir d'assemblages de négatifs photographiques, ce que son collègue Henry Peach Robinson appela plus tard *combination printing*. Mais elle prend une dimension nouvelle, par sa massification et par un phénomène qui est lié à la méthode de production de ces images : la disparition de la référence indiciaire même la plus élémentaire. Dans le photomontage historique, on reconnaît tel personnage, tel objet, bref un référent, aussi fragmentaire soit-il. Or avec la production d'images par des algorithmes, la machine intervient au niveau du pixel, de sorte qu'elle ne compose pas à partir de fragments, mais à partir d'un calcul de probabilité. Rien dans l'image ne fait référence à quoi que ce soit de réel.

AS Je suis tout à fait d'accord sur le fait que nous vivons depuis quelques années un tournant majeur dans notre rapport aux images. Ce tournant est causé en grande partie par le développement de plus en plus rapide des technologies de *deep learning* (ou d'« apprentissage profond ») ayant recours à des réseaux de neurones artificiels et profitant de la possibilité d'accéder à de vastes bases de données d'images composées à partir des milliards d'images numériques accessibles en ligne. Ce tournant majeur dans la culture visuelle contemporaine – dont la complexité et les enjeux sont comparables à ce qui s'est passé au début des années 1990, avec la diffusion des technologies numériques appliquées aux images – se manifeste aujourd'hui, il me semble, à travers trois phénomènes étroitement liés les uns aux autres.

Tout d'abord, l'utilisation de processus de *deep learning* pour activer des systèmes de *machine vision* (« vision artificielle » ou « vision par ordinateur »). Ceux-ci peuvent être appliqués à n'importe quelle image numérique accessible en ligne : des images qui sont *machine-readable* (c'est-à-dire susceptibles d'être analysées de manière automatique afin d'en extraire des informations et des données) et qui le sont *même quand elles ne sont pas visibles par des yeux humains*.

Ensuite, l'utilisation de processus de *deep learning* pour *transformer* des images existantes, à travers plusieurs types d'opérations, ou pour *générer* des images entièrement nouvelles, ayant tantôt un aspect parfaitement mimétique et photoréaliste, tantôt un aspect hybride où elles restent reconnaissables dans une certaine mesure, tantôt encore un aspect entièrement abstrait.

Finalement, l'utilisation de processus de *deep learning* pour établir de nouvelles relations entre mots et images : d'un côté, en produisant de manière automatique des images à partir de mots, phrases ou textes (comme cela se passe avec des logiciels et applications comme DALL-E ou Craiyon) ; de l'autre, en générant, toujours de manière automatique, des mots, phrases ou textes qui « décrivent » ce qui est représenté dans une image, avec une forme de légendage automatisé ou *neural caption* qui peut être presque considérée comme une nouvelle forme d'*ekphrasis*. On peut voir ces deux opérations à l'œuvre, avec une forme d'enchaînement, dans un projet comme *Contrefaits* (2022) de Grégory Chatonsky, dans lequel on part d'une image provenant de l'histoire de l'art, comme par exemple *Les Demoiselles d'Avignon* (1906-1907) de Picasso ; on demande à un logiciel de la décrire, avec une *neural caption*, et on utilise ensuite le texte généré automatiquement comme *prompt* pour produire une nouvelle image.

Les images hautement photoréalistes mais dépourvues de référents profilmiques réels – comme par exemple les portraits de la série *This Person Does Not Exist* (réalisée en 2019 avec StyleGAN2, un Generative Adversarial Network introduit par des chercheurs de Nvidia en 2018) – font donc partie d'un vaste spectre de phénomènes dans lesquels se manifeste aujourd'hui l'impact de l'intelligence artificielle (plus spécifiquement : des techniques de *deep learning*) sur la culture visuelle contemporaine.

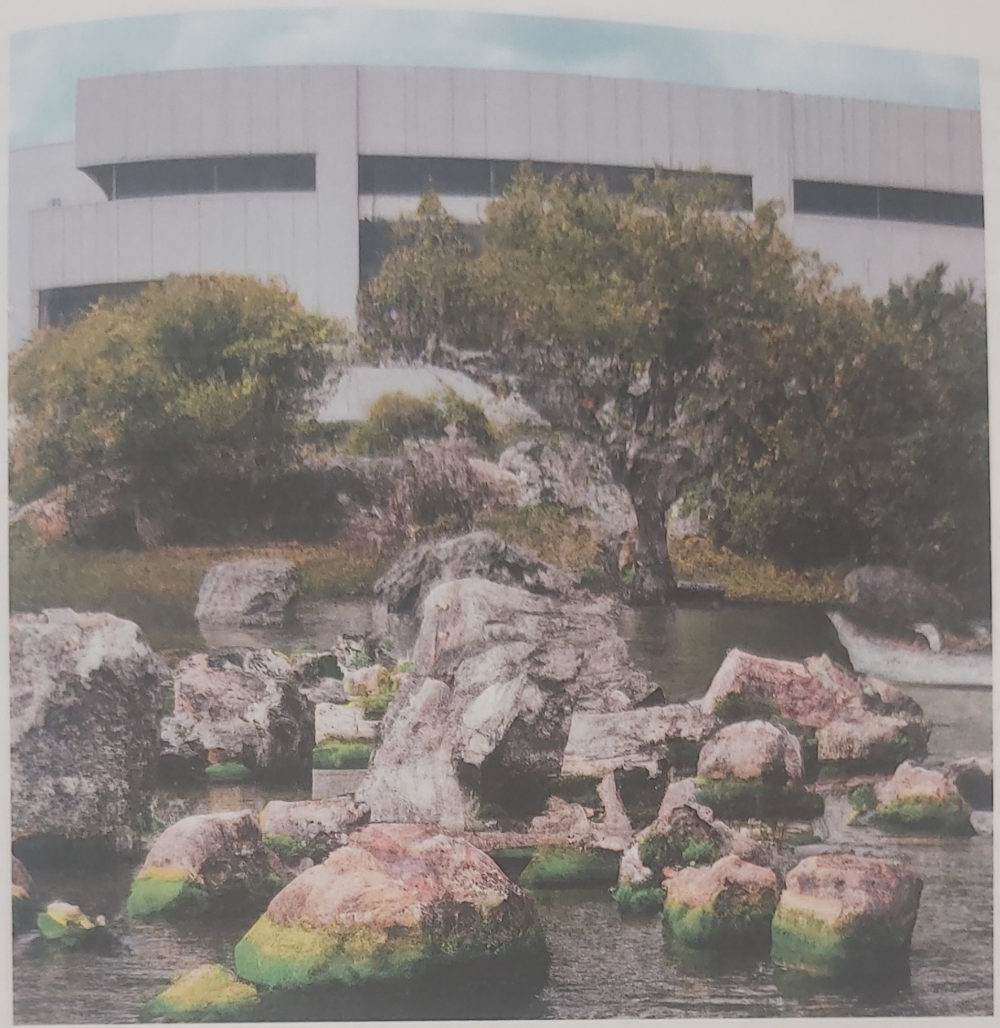
2. Grégory Chatonsky, *Contrefaits : Les Demoiselles d'Avignon*, 2021, version modifiée de Neural Story Teller et Disco Diffusion.

CJ La perte du référent, c'est ce que tu as appelé « disréalisme », Grégory, n'est-ce pas ?

GC L'idée d'un référent perdu n'est pas nouvelle. Elle semble n'avoir jamais existé qu'à travers une négativité, à savoir la biffure de la chose supposée représentée ou son retrait. Aussi, on a tendance à penser que ce que j'ai nommé ailleurs les « images d'images » nous font perdre ce que les photographies nous auraient offert depuis la révolution industrielle : l'indice d'une chose en soi. Or, il me semble préférable de suspendre nos croyances photo-ontologiques afin d'y voir l'émergence d'une nouvelle forme de réalisme ou de mimétisme, et d'engager un travail de reconstruction historique qui met cette évolution en perspective : la traduction de toutes choses en énergie *(energeia)*, c'est-à-dire en kilojoule, fournissant un travail thermodynamique et produisant de la valeur monétaire pour être converti en code binaire où le courant passe (1) ou ne passe pas (0). Chaque chose étant codée selon la même norme, permettant la totalisation du calcul.

Il s'agirait ainsi de reconstituer ce qu'on pourrait nommer la logistique des images qui les fait circuler de plus en plus autour de la Terre. Il faudrait pour cela tenir compte de leur reproductibilité technique – à savoir leur support matériel (extraction) –, de la massification de leur usage de capture et penser cette logistique jusqu'à l'hypermnésie du Web. Tous ces aspects ont transformé jusqu'à l'historicité même des images. Ce n'est pas une histoire parmi d'autres, car cette nouvelle modalité d'existence des images a modifié les conditions de possibilité mêmes de l'histoire. Comment les historiens futurs, à supposer qu'il y en ait, pourront-ils narrer notre époque où la discrimination entre le mémorable et l'oubliable, l'œuvre et le commun s'est effondrée ? Que feront-ils de tous ces documents dont la quantité remet en cause la notion même d'archive ? Quelle odeur historique pourront-ils bien humer ? Il y a dans les données massives du réseau quelque chose qui n'est pas accessible à un système nerveux humain, une quantité qui nous exproprie de notre propre mémoire. Nous avons déposé sur le Web des traces existentielles en un si grand nombre que le temps exigé pour les consulter engorgerait le présent. À qui ou à quoi sont-elles destinées ?

Il faut peut-être ouvrir la possibilité d'un changement radical du « pacte indiciaire », car ces images générées par le traitement statistique de millions d'images sont bien reconnaissables, même si elles n'ont jamais existé. Voilà le paradoxe, la reconnaissance mimétique, la représentation n'est pas dépendante d'un lien univoque avec ce qui est. Elle est aussi fonction de ce qui est possible, c'est-à-dire de ce qui pourrait exister pour le destinataire. L'induction



3. Grégory Chatonsky, *Les Images de Morel*, 2022, adaptation de l'invention de Morel d'Adolfo Bioy Casares avec DALL-E 2.

statistique n'est pas un collage, qui est d'une certaine façon la forme visuelle de la citation textuelle, ce n'est pas un fragment sélectionné puis retravaillé pour constituer un objet de culture, c'est une synthèse inductive : partant par exemple de milliers d'images d'oiseaux et calculant la fréquence des pixels, le logiciel peut produire des assemblages de pixels (un bruit plus ou moins organisé, plus ou moins abstrait) que nous interprétons comme des oiseaux sans qu'il y ait dans le logiciel la moindre idéation du concept d'oiseau. Le réseau de neurones est simplement capable de discriminer entre les résultats, de choisir le possible ou le vraisemblable et de limiter la marge d'erreur. Les images générées pourront-elles être réinjectées dans le stock d'apprentissage ? Alors, l'extension prendrait le pas sur la définition. Fin des Formes Idéales, donc, et du mythe de la réflexivité au profit de la récursivité.

La notion de disréalisme tente d'approcher ce paradoxal réalisme des possibles. Le préfixe « dis » introduit une séparation, une différence, une cessation ou un défaut dans la réalité. « Il y a un autre monde, mais il est dans celui-ci », écrivait Ignaz Paul Vital Troxler. L'*imagination* artificielle des réseaux de neurones – que je préfère à la notion d'*intelligence* artificielle – ouvre sur ce défaut au cœur de la réalité parce que, si ces images ne sont pas réelles, elles existent bel et bien et appartiennent dès lors au stock d'images disponibles. On pourrait proposer le concept de « dispossible » pour souligner la conjonction du disponible et du possible. Il faut préciser que cette fissure de la réalité est généralisée, puisque l'induction statistique s'applique à beaucoup de domaines de connaissance dans les sciences dures et humaines. Comme le fait le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) dans le domaine de la prévision du changement climatique, nous travaillons sur des scénarios multiples et probables auxquels nous accordons plus ou moins de crédibilité selon nos systèmes de croyance et l'idée que nous nous faisons des croyances des autres[1].

CJ Cet état de fait est-il pour autant radicalement nouveau ?

AS À mon avis, on ne peut pas parler d'une nouveauté « radicale », c'est-à-dire produisant une fracture nette, un moment de discontinuité totale par rapport à ce qui a précédé l'arrivée des technologies de *deep learning*. Quand on étudie l'histoire des médias, il faut toujours essayer d'identifier autant les moments de rupture que les lignes de continuité, et les deux sont souvent entremêlés. Chaque nouveau médium est aussi une forme de *remédiation* (je cite le concept formulé par Jay David Bolter et Richard Grusin[2], qui à son

tour est une reformulation d'une idée de Marshall McLuhan) de certains médias qui l'ont précédé. Selon cette perspective, chacun des trois phénomènes que j'ai mis en évidence peut et doit être situé dans une perspective historique, afin de comprendre les lignes généalogiques – souvent multiples et croisées – dans lesquelles il s'insère. Les systèmes de *machine vision*, par exemple, peuvent être mis en relation avec l'histoire des tentatives d'« automatiser » la perception visuelle, et avec celle de l'idée même d'un « regard de la machine » : une idée qui traverse l'histoire des théories de la photographie et du cinéma, et qui, dans le temps, a soulevé toute une série d'espoirs et de craintes liées à la possibilité d'un dépassement ou d'un décentrement du regard humain.

Pour ce qui concerne les images photoréalistes générées par des technologies de *deep learning* comme les Generative Adversarial Networks (GAN) ou « réseaux antagonistes génératifs », elles peuvent en effet être situées dans l'histoire des « images d'images », vu qu'elles sont générées à partir d'un processus d'induction statistique, au moins en partie automatique, qui se fonde sur l'analyse préalable de vastes bases de données d'images. Le problème, comme le souligne justement Grégory, est de comprendre avec quelles « images d'images » on peut les mettre en relation. Les compositions allégoriques à partir d'assemblages de négatifs photo d'Oscar Rejlander, les doubles tirages des « marines » de Gustave Le Gray, les portraits composites de Francis Galton, les surimpressions de la photographie spirite, et ensuite l'essor du photocollage, du photomontage et des surimpressions dans la photographie amateur des années 1890-1910[3], puis dans la photographie des avant-gardes des années 1920, font partie de cette histoire, mais ce ne sont pas nécessairement les formes et les techniques qui nous permettent de mieux comprendre les nouvelles « images d'images » générées par l'IA.

Comme Grégory le souligne justement dans sa réponse, avec ces images, nous sortons d'une logique fondée sur l'agencement de fragments (photocollages et photomontages) et sur la superposition de couches (surimpressions) pour accéder à une logique fondée sur la réduction de toute image à une configuration de pixels disposés à travers une grille orthogonale, et sur la possibilité de reproduire cette configuration (ou une configuration semblable) à travers une forme d'induction statistique. Comme dans le cas de toute image numérique, nous sommes donc pleinement situés dans la ligne généalogique des images *matricielles*, organisées en forme de grille de points (de tissu, d'encre, de lumière), mais à cette organisation réticulaire et cartésienne s'ajoute une autre couche, qui est celle du calcul statistique alimenté par l'intelligence artificielle.

CJ Il y a ensuite la question du « réalisme » des images photoréalistes générées par des technologies de *deep learning*. Comment le qualifier ?

AS La notion de « disréalisme » proposée par Grégory me semble tout à fait efficace, parce qu'elle garde l'idée qu'il y a là une nouvelle forme de « réalisme », tout en soulignant, avec le préfixe « dis- », sa différence profonde par rapport au « réalisme » auquel on a historiquement associé la photographie argentique (considérée en tant que trace ou empreinte sur une surface photosensible) et auquel on continue aussi à associer une vaste quantité de photographies ou de *stills* vidéos numériques, malgré l'abandon du paradigme de la trace et de l'empreinte.

À cette notion de « disréalisme », j'aimerais quand même ajouter la tentative de mieux comprendre celles qui à mon avis sont de nouvelles formes d'« indicialité » ou de « référentialité » spécifiques des images photoréalistes produites par le *deep learning*. À bien voir, en effet, ces images ne sont pas complètement dépourvues de référent. Au contraire, elles semblent être caractérisées par la présence de plusieurs « couches d'indicialité » ou de « référentialité ». Elles sont générées par une forme d'induction statistique automatisée par le *deep learning*, mais cette induction se base sur l'analyse préalable de vastes bases de données d'images, et ces images, prises dans leur ensemble, sont un premier référent.

Ensuite, dans la grande majorité de ces images qui font partie des bases de données, nous voyons représentées des entités (objets, corps, visages, gestes, expressions, lieux, etc.) qui à un moment ou un autre ont été captées par une caméra : il s'agit d'entités « profilmiques », qui constituent un deuxième niveau de référentialité. Ensuite encore, il y a tous les mots et toutes les phrases qui ont été utilisés pour « indexer » ou « étiqueter » ces images dans les bases de données (dans une base de données souvent utilisée comme ImageNet, par exemple, il s'agit de 14 millions d'images organisées en 21 000 catégories et sous-catégories basées sur la structure sémantique de Wordnet, une base de données de classifications de mots développée à Princeton dans les années 1980) : l'ensemble de ces mots et de ces phrases, de ces légendes *(captions, tags)* constitue encore un autre niveau de référentialité.

Finalement, on pourrait dire que, comme toute image matérielle (c'est-à-dire pas seulement mentale), les images photoréalistes générées par le *deep learning* renvoient aussi aux algorithmes à travers lesquels elles ont été générées, comme une photographie argentique renvoie à l'ensemble des appareils optiques, des supports photosensibles, des processus chimiques et des techniques d'impression qui ont été mobilisés pour la produire. Les images générées par des Generative Adversarial Networks, par exemple, quand elles ne sont pas parfaitement photoréalistes mais plutôt hybrides ou abstraites, sont assez facilement identifiables, comme s'il y avait un « style » propre aux images produites par cette classe d'algorithmes et de neurones artificiels. Certains commentateurs[4] ont d'ailleurs formulé le terme « GANisme » pour donner un nom à ce type de « style », dont on peut trouver les traces dans les œuvres de plusieurs artistes contemporains qui travaillent – souvent, de manière très différente, comme on peut le voir en prenant en considération des artistes comme Grégory Chatonsky, Trevor Paglen, Hito Steyerl, Pierre Huyghe, et beaucoup d'autres – avec ces technologies.

4. Grégory Chatonsky, *La Machine 100 têtes : Ce singe, serait-il catholique par hasard ?*, 2022, adaptation de *La Femme 100 têtes* (1929) de Max Ernst avec DALL-E 2.

CJ Comme vous venez tous les deux de le dire, toutes les technologies automatisées dont nous parlons ici travaillent à partir de *deep learning*, autrement dit, elles dépendent de calculs de probabilités et de statistiques établies à partir de gigantesques bases de données comme ImageNet. Elles ne font donc que rarement appel à de la modélisation, telle qu'on peut la pratiquer en architecture ou dans la production d'effets spéciaux au cinéma. Ce qui nous amène à penser la production d'images composites sous l'angle d'une forme algébrique plutôt que géométrique, alors même que l'histoire de l'art du XX^e^ siècle avait valorisé une modernité géométrique à travers l'artiste-ingénieur, notamment dans le constructivisme des années 1920. L'intelligence artificielle nous projette dans le monde de la probabilité, de l'algèbre, de la statistique, n'est-ce pas ?

GC En effet, il existe plusieurs catégories d'images générées. La première, qui a été populaire pendant des années, c'est l'image de synthèse largement utilisée pour les FX, les effets spéciaux. Ce sont des modélisations scientifiques. On incorpore dans des logiciels les connaissances hypothético-déductives du monde physique que nous avons : gravité, résistance, turbulences, etc. Ce sont des images laplaciennes qui supposent que si nous connaissions l'ensemble des lois scientifiques qui régissent l'univers, nous pourrions connaître l'état de cet univers dans l'avenir et, d'un point de vue visuel, mimer parfaitement la réalité (ce fut le fantasme de la réalité virtuelle dans les années 1980-1990). C'est le désir d'une fusion entre la connaissance (scientifique) et les conditions d'expérience.

La seconde est l'induction statistique : on entraîne un logiciel sur un très grand nombre de données binaires et il va calculer des statistiques, la probabilité pour que telle ou telle unité, un pixel par exemple, apparaisse afin que le bruit visuel converge vers quelque chose que nous pouvons reconnaître. C'est ce qu'on appelle l'« espace latent », un espace virtuel où sont conçues, accumulées et distribuées les images-possibles comme données numériques et outils de calcul. Cet espace latent est assez difficile à conceptualiser tant sa généralité est abstraite. Les images n'y sont plus des images, mais sont décomposées sous forme de 0 et 1. Si nous reprenons notre exemple des oiseaux, ceux-ci vont avoir une forme qui se détache d'un arrière-plan simplement parce que dans le stock d'images qui nourrit le réseau de neurones, la probabilité pour que les pixels fassent apparaître une telle séparation est proche et se traduit en vecteurs, c'est-à-dire un ensemble de points orientés. De sorte qu'il faut imaginer l'espace latent comme des points disposés dans un espace dont les regroupements constituent par exemple les catégories (nommées aussi paramètres) des étants « oiseau », « chat », « lion » ou des catégories visuelles « rouge », « vert », « bleu », ou encore « flou », « net », « proche », « loin ». Il n'y a aucune déduction idéative dans cette catégorisation, mais simplement une classification déduite de la probabilité d'apparition de chaque pixel corrélé à une nomenclature textuelle *(tags)* faite par des humains image par image et contextualisée par Wordnet. C'est le modèle CLIP qu'on retrouve dans DALL-E ou Stable Diffusion. Quand on génère une image d'oiseau, on reste dans l'espace latent de l'oiseau, mais

on peut tout aussi bien créer un hybride entre un animal et un instrument, entre la technique et l'organique en pliant l'un sur l'autre ces deux espaces distants et en produisant ainsi des proximités non réalistes. La manière dont on se déplace dans cet espace produit des métamorphoses, la façon dont on le plie génère des hybrides. Un espace latent n'est donc pas constitué d'images, mais contient des images-possibles en tant qu'elles sont décomposées selon leur unité, le pixel, et leur proximité statistique. Ces images-possibles sont les données passées du dataset (on peut faire le test de générer une image déjà existante afin de vérifier que la mise en statistique est complète) qui ont servi à l'entraînement, mais aussi les images futures qui n'existent pas encore. Un espace latent est structurellement ambivalent, car il peut servir à surveiller, c'est-à-dire à identifier la régularité statistique, ou à générer, c'est-à-dire à multiplier les images[5]. L'image-possible transforme la relation entre le passé et le futur et bouleverse le pacte indiciaire comme l'ayant-été photographique pour former une structure hyperstitionnelle : nous sommes hantés par les images-possibles du futur après l'avoir été par celles du passé en raison de l'hypermnésie existentielle du Web.

CJ Comment sont composées ces images ?

GC Leur compositionnalité est singulière. Elle ne fonctionne pas par prélèvement, fragmentation, coupure choisie, mais par induction automatique : toute image, puisqu'elle est binaire, a été codée en 0 et 1 et se retrouve dans l'espace latent, numériquement égale à toutes les autres images, selon une ontologie plate au sens de l'ontologie orientée, objet dont parle Graham Harman[6]. Cette mise à plat permet la traduction d'une chose en autre chose (par exemple rien n'empêche de recoder une image en son), une compatibilité générale, mais aussi la transduction, un changement de forme qui garde des traces de la phase précédente : il est possible de se déplacer dans l'espace latent et de passer d'une forme à une autre selon une lente métamorphose qui produit des étapes intermédiaires où deux paramètres sont co-présents[7]. L'air de famille, la paréidolie et le déjà-vu se généralisent.

Le pacte a donc changé de nature. Nous ne reconnaissons pas telle ou telle figure se découpant sur un fond, selon un *ethos* de la « postproduction » qu'évoque Nicolas Bourriaud dans son livre[8], où des éléments saillants de la

5. Grégory Chatonsky, *Laocoon VI*, 2022, image générée par DALL-E 2.

culture, tel ou tel motif, telle ou telle œuvre connue, pourraient être scénarisés à nouveau, déplacés et recontextualisés. Nous reconnaissons la platitude des images, c'est-à-dire leur binarité qui a permis leur calculabilité et leur inclusion dans l'espace latent produisant une ressemblance dépassant l'individualité de l'image pour se répandre entre chacune d'entre elles : dans l'espace latent il n'y a plus d'images, au sens strict du terme, mais des probabilités statistiques qui peuvent s'actualiser en débouchant sur des résultats visuels différents. En passant d'une image à une autre, on forme la traîne d'une métamorphose dans l'espace latent, on passe d'un ensemble de probabilités à un autre ensemble.

L'ontologie plate de l'espace latent n'est pas seulement une condition de production, elle peut être aussi considérée comme une représentation de la culture elle-même, à la suite d'un long processus d'industrialisation, dans sa forme la plus indéterminée possible : toutes les images passées, toutes les images futures. Car cette technologie ne produit pas seulement des images, elle cartographie en retour, si on l'interprète en replaçant ses résultats dans le champ culturel, l'imaginaire d'une époque, en nous demandant d'imaginer quelles propositions textuelles nous voulons former en vue de produire des images. En récupérant les millions de *prompts* déjà existants, on pourrait dresser cet autre espace latent culturel de l'imaginaire[9]. Au regard de l'histoire des relations entre le texte et l'image, cette nouvelle configuration soulève de nombreuses questions.

CJ On a beaucoup parlé d'induction, de statistique et de calcul de moyennes. Or le calcul de moyennes, c'est justement la principale activité de Francis Galton, l'inventeur de la photographie composite, cousin de Charles Darwin et anthropologue. Son grand projet est de produire des mesures de l'intelligence humaine, il est l'inventeur de la notion de quotient intellectuel – notion mise à mal depuis, notamment par Stephen J. Gould[10] – et de l'eugénisme. Or son ombre funeste plane sur beaucoup de technologies récentes, notamment Eigenface.

AS Le terme Eigenface indique une série de technologies de vision par ordinateur qui, à partir du début des années 1990, a permis le développement du champ de la reconnaissance faciale. Connue aussi par le terme de Principal

6. Des Eigenfaces provenant des AT&T Laboratories, Cambridge, 2022, images générées par le logiciel.

Component Analysis (PCA), la méthode des Eigenfaces (terme provenant de « eigen », qui veut dire « propre » ou « spécifique » en allemand, et « face »), permet de réduire chaque visage à un nombre fini de caractéristiques spécifiques, en réduisant donc aussi la quantité de données qu'on doit traiter pour détecter et reconnaître un visage.

L'idée d'établir une comparaison entre les *composite portraits* de Galton et les Eigenfaces m'est venue en regardant la manière dont les Eigenfaces sont visualisés. Comme dans le cas des *composite portraits*, ce que nous voyons, ce sont des visages flous, à peine reconnaissables. Et pourtant, *composite portraits* et Eigenfaces poursuivent des objectifs presque opposés. Dans le cas des *composite portraits*, le procédé technique de la surimpression est utilisé pour *superposer* plusieurs portraits photographiques d'individus appartenant à ce que Galton considérait comme une même classe ou une même partie de la société, afin d'individuer un visage « moyen ». Dans ce cas, le flou généré par la surimpression de plusieurs négatifs pouvait être considéré comme une analogie visuelle du calcul statistique des moyennes. Dans le cas des Eigenfaces, au contraire, le visage flou que nous voyons contient seulement les traits qui sont spécifiques à *un seul visage*, en somme *uniques*. Le flou, ici, n'est pas le résultat de la disparition de ce qui est *spécifique* aux différents visages, comme dans le cas des *composite portraits* de Galton, mais plutôt de la disparition de ce qui est *commun* à différents visages. Ce qui néanmoins relie les deux phénomènes – *composite portraits* et Eigenface – est l'application d'une logique statistique à l'analyse du visage, avec une finalité d'identification : identification d'un « type » social, dans le cas de Galton ; identification d'un individu, dans le cas des Eigenfaces.

CJ Antonio, depuis quelque temps tu travailles sur les enjeux de la haute et de la basse définition des images, une question qui se pose aussi dans le cas de la technique du *upscaling*, autrement dit la réintroduction d'informations dans les « images pauvres », selon l'expression d'Hito Steyerl, c'est-à-dire de basse définition ou de mauvaise qualité. Il me semble que ça touche très précisément notre question sur les images composites à l'ère des intelligences artificielles.

AS Mon intérêt pour le phénomène du *upscaling* – l'augmentation du degré de définition d'une image fixe ou en mouvement (et éventuellement de son *frame rate*, dans le cas des images en mouvement) – s'insère en effet dans la lignée de mes recherches sur les enjeux de la haute et de la basse définition des images. Il est aussi lié à ma tentative de dresser une liste des *opérations d'images* qui se manifestent dans les trois grands champs que j'ai mentionnés dans ma première réponse : les systèmes de *machine vision* appliqués aux milliards de *machine-readable images*, les images *transformées* ou *générées* par des systèmes de *deep learning*, les nouvelles formes d'automatisation de la mise en relation de textes et d'images.

La distinction entre haute et basse définition a des enjeux très importants d'ordre non seulement technologique, mais aussi esthétique, épistémologique, économique, politique et historique, comme nous avons essayé de le montrer, Francesco Casetti et moi, dans le livre collectif que nous avons dirigé avec le titre *La haute et la basse définition des images. Photographie, cinéma, art contemporain, culture visuelle*[11] et dans le numéro spécial de la revue en ligne *Necsus* sur le thème « Resolution »[12]. Les enjeux historiques sont particulièrement intéressants, parce que les signes de la matérialité du support (grains, rayures, altérations chromatiques) et le degré de définition plus ou moins élevé d'une image numérique peuvent être interprétés comme une marque de son statut temporel : la trace visible de son appartenance à une phase précise de l'histoire des technologies visuelles, argentiques ou numériques.

Or, dans le cas du *upscaling*, nous faisons face à des technologies encore une fois fondées sur le *deep learning* et sur l'accès à de vastes bases de données d'images qui permettent d'augmenter le degré de définition d'une image ainsi que le *frame rate* (nombre d'images par seconde, qu'on pourrait considérer comme une forme de définition « temporelle »), avec, pour conséquence, l'altération des marques qui caractérisent une image en tant que document historique.

Un des exemples les plus frappants et les plus discutés de ce phénomène du *upscaling*, ce sont les versions de films Lumière qu'on peut désormais aisément trouver en ligne. Dans certaines de ces versions, un film iconique comme *Arrivée d'un train à La Ciotat* (1897, Vue Lumière n° 653), grâce au processus du *upscaling*, a été passé d'une vitesse de défilement de 16 images par seconde à une vitesse de 60 *frames* par seconde, du format original 1.33:1 au format 16:9 typique des écrans d'ordinateurs contemporains, et du grain et des rayures d'une pellicule 35 mm de 1897 à une très haute définition numérique en 4K. À ces changements, dans l'une des versions de la vidéo accessibles en ligne[13], s'ajoute la colorisation d'images qui à l'origine étaient en noir et blanc[14] : cette opération avait déjà suscité plusieurs interrogations lors de la diffusion des documentaires de la série *Apocalypse*[15].

Qu'est-ce qui se passe quand une simple recherche sur Google ou YouTube du film Lumière nous fait tomber tout d'abord sur la version soumise au *upscaling*, et seulement après sur une version numérique du format argentique original ? Et si cette opération du *upscaling* des documents visuels historiques devait continuer à se diffuser, de quelle manière transformerait-elle notre relation avec l'histoire, avec les traces visuelles du passé ?

Ce qui rend aussi le phénomène du *upscaling* très intéressant est qu'il fait partie d'un vaste éventail d'*opérations* avec lesquelles les processus de *deep learning* sont en train de bouleverser profondément la culture visuelle contemporaine. Avec le *upscaling*, on fait une opération d'*augmentation* du niveau de définition d'images fixes ou en mouvement, qui a pour conséquence d'altérer les marques temporelles de ces images. Avec d'autres applications du *deep learning* au champ des images numériques, on peut effectuer d'autres opérations comme détecter, identifier, reconnaître, contrôler, monitorer, surveiller, réparer, complé-

ter, animer, rajeunir, ressusciter, imiter, transférer (un style, d'une image à une autre), décrire, visualiser, simuler, prédire, etc. Avec Ada Ackerman et Alice Leroy, nous avons essayé de dresser une cartographie de ces opérations alimentées par le *deep learning* dans un volume à paraître en 2023[16]. Dans notre introduction, nous soulignons le fait que les applications du *deep learning* à la culture visuelle contemporaine nous obligent à revenir sur l'ancienne question du pouvoir, de la puissance et de l'agentivité des images sur laquelle se sont penchés, depuis la fin des années 1980, des historiens de l'art, des théoriciens des images et de la culture visuelle et des anthropologues comme (parmi beaucoup d'autres) David Freedberg, Hubert Damisch, Georges Didi-Huberman, Alfred Gell, W.J.T. Mitchell, Horst Bredekamp et Philippe Descola. Cette question montre aussi que le concept d'« image opératoire » formulé par Harun Farocki au début des années 2000, en lien avec des vidéoinstallations comme *Eye/Machine I*, *II* et *III* (2001-2003) est encore tout à fait… opératoire, et nous invite à détecter et répertorier les « opérations d'images » dans lesquelles interviennent des technologies de *deep learning*.

GC Nous nous sommes immédiatement appelés et avons partagés notre émotion avec Antonio, la première fois que nous avons vu cette séquence de La Ciotat alors que certains de nos camarades dénonçaient une « trahison ». Car si elle est upscalée, c'est au sens d'une complétion : grâce à un dataset d'images prises entre 1895 et aujourd'hui, le logiciel propose de compléter les images, de les « réparer », d'en créer des versions intermédiaires. Cette complétion inductive est différente de la colorisation à la main, de la même manière que sont différentes la synthèse inductive et la citation ou le découpage visuel. En effet, ici c'est la mémoire statistique des images nous séparant de l'original qui permet de la compléter et d'halluciner des détails dans le bruit de l'image. C'est donc un profond changement du statut même de l'archive : le photoréalisme est alors conquis par une logique qui le dépasse. L'indice de ce qui a été à un moment précis devient la traînée d'un temps plus long au cours duquel le dataset a été constitué. Ce sont toutes les images passées qui réparent ces images comme si elles étaient solidaires les unes des autres, comme si l'agentivité entre elles ouvrait la possibilité d'une autre histoire. On peut

7. Grégory Chatonsky, *Les Vues imaginaires : Aubervilliers*, 2022, archives modifiées de la ville d'Aubervilliers par DALL-E 2.

alors créer des versions contrefactuelles de documents existants, les étendre, les compléter, les réparer. Cette déconstruction de l'archive est une hantologie : lorsque nous regardons l'entrée en gare de La Ciotat, le grain de l'image n'appartient plus à la fin du XIX[e] siècle, mais au Portapak Sony de la fin des années 1960. L'espace latent hante à présent toutes les images.

CJ Nous savons qu'à ce jour les algorithmes ne définissent pas les formes qu'ils produisent. Grégory l'a bien expliqué. À aucun moment ces calculs de probabilité ne tiennent compte des formes qui émergent, ne les reconnaissent comme formes émergentes, car ils n'opèrent que par des calculs de probabilité statistiques appliqués pixel par pixel. Il en va de même des sujets des images. Si j'ai bien compris le fonctionnement de DALL-E, l'algorithme ne connaît les sujets que parce que les images sont indexées et non par reconnaissance visuelle. Il n'analyse pas davantage la syntaxe des phrases qu'on leur donne en aliment, si bien qu'ils n'ont aucune intelligence de la composition des images. Or, nous reconnaissons *in fine* des formes que nous arrivons à nommer : un visage, une montagne, un bâtiment sur tel ou tel fond. Il y a quelque chose de stupéfiant dans cette rencontre entre le pur calcul de probabilité sans intention formelle et la reconnaissance par le spectateur de formes identifiables.

GC Il me semble que l'automatisation de la compositionnalité a en effet des résultats étonnants qui ne devraient logiquement pas avoir lieu sans idéation. C'est pourquoi il est nécessaire de décorréler les causes et les effets technologiques : des causes réductionnistes peuvent produire des effets complexes dans la mesure où ces effets sont hétéronomes en tant qu'ils rencontrent des agents humains dans un milieu associé. Mais ces résultats mettent aussi en doute un certain mythe de la réflexivité humaine : si des processus aussi simples peuvent produire un monde de sens, sommes-nous bien sûr que l'idéation fonctionne comme nous le croyons ? C'est en particulier la cohérence géométrique, lumineuse et stylistique qui est étonnante, puisqu'elle n'a pas été modélisée mais paramétrée. Il s'agit de comprendre qu'à partir d'un grand nombre d'images, on parvient à faire démêler tous ces paramètres et qu'on associe dans un *prompt* une description des étants de l'image souhaitée et une description technique des objectifs utilisés, des appareils, des pellicules, des styles, etc., parce que celle-ci est présente automatiquement dans la plupart des métadonnées des images. Il y a ainsi dans ces images quelque chose de kitsch. Lorsque nous écrivons un *prompt*, nous projetons une image qui n'existe pas encore et qui pourrait être. Ce pouvoir-être de l'image est la corrélation entre l'espace latent de l'imagination artificielle et l'espace latent culturel d'interprétation, corrélation qui prend la forme de l'adaptation de notre manière d'écrire aux capacités d'un logiciel.

L'ensemble des paramètres, leur traversée et leur pliage permettent de produire des images réalistes parce qu'elles sont nimbées de l'espace latent qui porte un air de famille généralisé. Ils permettent aussi de poursuivre une image au-delà de son cadre en mimant son style, sa géométrie, sa texture, etc. Ainsi les documents d'archives se transforment en possibilité d'autres images, la production ne s'arrête pas à la « prise » qui n'en est qu'une des étapes : nous poursuivons les images. Après l'accumulation hypermnésique du Web, leur numérisation et leur catégorisation en bases de données, leur mise en statistique, les images semblent intensifier une accélération qui se retourne sur nous-mêmes comme un gant pour nous prendre de vitesse.

CJ Vision vertigineuse et borgésienne…

8. Grégory Chatonsky, *Logistics : Dream Surface*, 2022, image générée par DALL-E 2 à partir d'un roman en cours d'écriture.

1 Anna Longo, *Le Jeu de l'induction. Automatisation de la connaissance et réflexion philosophique*, Sesto San Giovanni, Éditions Mimésis, 2022.

2 Jay David Bolter et Richard Arthur Grusin, *Remediation. Understanding New Media*, Cambridge/Londres, MIT Press, 2000.

3 Clément Chéroux, *Avant l'avant-garde. Du jeu en photographie, 1890-1940*, Paris, Éditions Textuel, 2015.

4 Voir par exemple Pau Waelder, «Beyond GANism: AI as Conceptual Art», *CIACMTL*, nº 3, 15 avril 2020 (en ligne : ‹http://ciac.ca/en/ai-ciac-mtl-03-01/›).

5 Grégory Chatonsky, *Skin Print*, 2017 (en ligne : ‹http://chatonsky.net/skin-print/›).

6 Graham Harman, *The Quadruple Object*, Winchester, Zero Books, 2011. Voir aussi Tristan Garcia, *Forme et objet. Un traité des choses*, Paris, PUF, 2011. On trouvera entre ces ontologies plates considérant toutes choses de façon égale et ôtant toute détermination, toute intensité, et l'espace latent de l'IA une troublante ressemblance malgré les différences de moyens utilisés.

7 Grégory Chatonsky, *Recursive Cinema*, 2022 (en ligne : ‹http://chatonsky.net/recursive-cinema/›).

8 Nicolas Bourriaud, *Postproduction. La culture comme scénario : comment l'art reprogramme le monde contemporain*, Dijon, Les presses du réel, 2009.

9 Grégory Chatonsky, *The Prompts Book*, 2022 (en ligne : ‹http://chatonsky.net/prompts/›).

10 Stephen Jay Gould, *La Mal-mesure de l'homme. L'intelligence sous la toise des savants*, trad. J. Chabert, Paris, Ramsay, 1983.

11 Francesco Casetti et Antonio Somaini (dir.), *La haute et la basse définition des images. Photographie, cinéma, art contemporain, culture visuelle*, Sesto San Giovanni, Éditions Mimésis, 2021.

12 Francesco Casetti et Antonio Somaini (dir.), «Resolution», numéro spécial de la revue en ligne *Necsus*, printemps 2018 ‹https://necsus-ejms.org/portfolio/spring-2018_resolution/›.

13 La version *upscaled* et colorisée du film Lumière *Arrivée d'un train à La Ciotat* (1897) est accessible sur la chaîne YouTube *Deoldify videos* : ‹https://www.youtube.com/watch?v=EqbOhqXHL7E›.

14 Peter Geimer, *Die Farben der Vergangenheit. Wie Geschichte zu Bildern wird*, Munich, C.H. Beck, 2022.

15 Voir Georges Didi-Huberman, «En mettre plein les yeux et rendre Apocalypse irregardable», *Libération*, 22 septembre 2009 (en ligne : ‹https://www.liberation.fr/ecrans/2009/09/22/en-mettre-plein-les-yeux-et-rendre-apocalypse-irregardable_952332/›).

16 Ada Ackerman, Alice Leroy et Antonio Somaini (dir.), *L'Image à l'épreuve des machines. Reconfigurations du visible*, coll. «Carnets du BAL», nº 12, à paraître en 2023.